(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Offenlegungsschrift**

(10) **DE 197 23 909 A1**

(51) Int. Cl.[6]:
G 11 B 20/18
// G06F 12/16,11/20

(21) Aktenzeichen: 197 23 909.9
(22) Anmeldetag: 6. 6. 97
(43) Offenlegungstag: 2. 4. 98

DE 197 23 909 A1

(30) Unionspriorität:
41479/96 21.09.96 KR

(71) Anmelder:
Samsung Electronics Co., Ltd., Suwon, Kyungki, KR

(74) Vertreter:
Grünecker, Kinkeldey, Stockmair & Schwanhäusser, Anwaltssozietät, 80538 München

(72) Erfinder:
Lee, Hae-Seung, Kyungki, KR

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Verfahren zum Erzielen einer geteilten Paritätsersatzplatte in einem RAID-Untersystem

(57) Ein Verfahren zum Verbessern der Fehlerfestigkeit und Leistungsfähigkeit eines RAID-Untersystems mit gespaltener Parität und Paritäts- und Ersatzplattenlaufwerk speichert in verteilter Weise Daten in einer Plattengruppe, die aus mehreren Plattenlaufwerken besteht, und führt einen Eingabe/Ausgabe-Betrieb parallel aus. Zu diesem Zweck werden die Plattengruppen mit wenigstens zwei Datenplattenlaufwerken zum Speichern von Daten, einem Ersatzplattenlaufwerk zur Verwendung bei Ausfall eines Plattenlaufwerks und ein Paritätsplattenlaufwerk zum Speichern von Paritätsdaten aufgebaut. Die Paritätsdaten des Paritätsplattenlaufwerks werden aufgeteilt, und die aufgeteilten Daten werden auf das Paritätsplattenlaufwerk und das Ersatzplattenlaufwerk aufgeteilt.

DE 197 23 909 A1

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

BUNDESDRUCKEREI 02. 98 802 014/650 27/22

Beschreibung

Hintergrund der Erfindung

1. Gebiet der Erfindung

Die vorliegende Erfindung bezieht sich auf ein Speichersystem großer Kapazität und insbesondere auf ein Verfahren zum Erzielen einer geteilten Paritätsersatzplatte zur Verbesserung der Fehlerfestigkeit und des Betriebs eines RAID-(Redundant Arrays of Inexpensive Disks = redundante Anordnungen aus billigen Platten) Untersystems.

2. Beschreibung des Standes der Technik

Die Leistungsfähigkeit eines Rechnersystems hängt von einer zentralen Prozessoreinheit und einem Eingabe/Ausgabe-Untersystem ab. In letzter Zeit führt die Entwicklung einer VLSI-(Hochintegrations-)Technik zu einer großen Verbesserung einer Arbeitsgeschwindigkeit der zentralen Prozessoreinheit. Da das Eingabe/Ausgabe-Untersystem jedoch nur eine leichte Leistungsverbesserung aufweist, ist das Verhältnis der Eingabe/Ausgabe-Verarbeitungszeit zur Gesamtverarbeitungszeit des Systems allmählich größer geworden. Weil außerdem eine allmähliche Steigerung der Datenwiedergewinnungskosten auf Auftreten von Fehlern im Eingabe/Ausgabe-Untersystem stattgefunden hat, besteht Bedarf für die Entwicklung eines Eingabe/Ausgabe-Untersystems, das gesteigerte Leistung und Zuverlässigkeit aufweist. Eine der Forschungen zur Verbesserung der Leistung des Eingabe/Ausgabe-Untersystems dreht sich um ein RAID-Untersystem. Ein allgemeines Eingabe/Ausgabe-Untersystem gibt Daten in ein Plattenlaufwerk sequentiell ein und davon aus, während das RAID-Untersystem einen Eingabe/Ausgabe-Betrieb parallel ausführt, indem Daten in einer Plattengruppe verteilt gespeichert werden, die aus mehreren Plattenlaufwerken besteht. Der Eingabe/Ausgabe-Betrieb wird dadurch schnell ausgeführt, selbst wenn ein Fehler auftritt, weil es möglich ist, Daten wiederzugewinnen, indem einfache Paritätsinformation verwendet wird. Damit ist auch die Zuverlässigkeit verbessert. Konkurrierend damit ist eine Technik, die sich auf das RAID-Untersystem bezieht, über den theoretischen Zustand hinaus eine kommerziell verwertbare Form entwickelt worden. In Universitäten sind aktive theoretische Studien unternommen worden über eine Studie am RAID-Algorithmus und ein Experiment unter Verwendung von Simulation. Unternehmen haben sich der Verbesserung der Eingabe/Ausgabe-Leistung und der Verbesserung der Zuverlässigkeit durch Ableitung von Dingen gewidmet, die durch zahlreiche Verbesserungsmaßnahmen reformiert werden sollen. Die Plattenanordnung ist in einem Superrechner, wie Cray, für die Eingabe/Ausgabe-Verbesserung des Plattenlaufwerks verwendet worden. Das Konzept des RAID wurde durch die Veröffentlichung durch drei Recherwissenschaftler der Berkeley-Universität 1988 geschaffen.

Fig. 1 zeigt ein typische RAID-Untersystem. Das RAID-Untersystem enthält einen Plattengruppensteuerer 4 und eine Plattengruppe 6 aus mehreren Plattenlaufwerken 6-0, ... 6-n. Der Plattengruppensteuerer 4, der zwischen ein Wirts-System 2 und die Plattengruppe 6 geschaltet ist, liest/schreibt Daten von/in die Plattenlaufwerke 6-0, ... 6-n der Plattengruppe 6 gemäß einem

Datenlese-/-schreibbefehl vom Wirts-System 2. In diesem Falle speichert der Plattengruppensteuerer 4 die Daten in verteilter Weise in jedes Plattenlaufwerk der Plattengruppe 6, verarbeitet einen Eingabe/Ausgabe-Betrieb parallel, und beim Auftreten eines Fehlers wiedergewinnt er die Daten durch Verwendung einfacher Paritätsinformation.

Fig. 2 zeigt den Plattengruppensteuerer 4 im Detail. Der Plattengruppensteuerer 4 enthält eine Wirts-Schnittstellensteuerer 10, einen Prozessor 12, einen Speicher 14, einen Pufferspeicher 16 und einen Gruppensteuerer 18. Der Wirts-Schnittstellensteuerer 10 vermittelt die gesendeten und empfangenen Daten zwischen dem Wirts-System 2 und dem Prozessor 12. Der Prozessor 12 steuert den Gesamtbetrieb des Plattengruppenuntersystems. Der Speicher 14, der mit dem Wirts-Schnittstellensteuerer 10 verbunden ist, hat einen Festspeicher ROM zur Speicherung eines Steuerprogramms des Prozessors 12 und einen Arbeitsspeicher (Speicher mit wahlfreiem Zugriff) RAM zum Zwischenspeichern von Daten, die während eines Steuervorgangs erzeugt werden. Der Pufferspeicher 16 speichert vorübergehend einen Datenlese-/Regenerierbefehl und Daten, die zwischen dem Wirts-System 2 und der Plattengruppe 6 gesendet und empfangen werden, unter der Steuerung des Prozessors 12. Der Gruppensteuerer 18 vermittelt und steuert zahlreiche Daten, die zwischen dem Prozessor 12 und der Plattengruppe 6 gesendet und empfangen werden.

Das in der oben beschriebenen Art aufgebaute RAID-Untersystem dient der Verbesserung der Leistung einer Eingabe/Ausgabe-Vorrichtung, vergrößert deren Kapazität und schafft deren Zuverlässigkeit durch Dezentralisierung oder Streifenzerlegung von Daten zum Plattenlaufwerk, die Spiegelung der Platte mit wiederholten Daten usw. Die RAID-Theorie ist auf eine Vorrichtung mit sequentiellem Zugriffs, wie beispielsweise eine Bandkassette, als Eingabe/Ausgabe-Vorrichtung anwendbar, jedoch ist ihre Hauptanwendung bei Festplatten.

Eine RAID-Struktur ist in sechs RAID-Niveaus vom Niveau 0 zum Niveau 5 entsprechend ihrer Eigenschaften klassifiziert. Die sechs RAID-Niveaus haben Vorteile und Nachteile, je nach Umgebung, die für jede Charakteristik geeignet sind, und werden in zahlreichen Anwendungsgebieten eingesetzt. Jedes RAID-Niveau schafft eine Lösung für zahlreiche Datenspeichervorrichtungen oder ein Zuverlässigkeitsproblem.

Der Inhalt jedes RAID-Niveaus wird nun beschrieben.

RAID-Niveau 0

Das RAID-Niveau 0 ist bei der Verarbeitung von Interesse anstatt der Zuverlässigkeit der Daten. Die Daten werden verteilt in allen Datenlaufwerken der Datengruppe gespeichert. Unterschiedliche Steuerer werden dazu verwendet, die Plattenlaufwerke der Plattengruppe miteinander zu verbinden. Das RAID-Niveau 0 hat einen Vorteil dahingehend, daß die Eingabe/Ausgabe-Leistung durch gleichzeitigen Datenzugriff wegen der Verwendung verschiedener Steuerer verbessert ist.

RAID-Niveau 1

Die Inhalte aller Plattenlaufwerke werden identisch in einem Kopieplattenlaufwerk gespeichert. Ein solches Verfahren nennt man Spiegelung. Das Spiegelungsver-

fahren verbessert die Leistung des Plattenlaufwerks, hat jedoch einen wirtschaftlichen Nachteil. Das RAID-Niveau 1 hat nämlich den Nachteil, daß nur 50% der Platte in einem System verwendet werden, das eine Plattengröße großer Kapazität verlangt, wie beispielsweise einem Datenbanksystem. Da die gleichen Daten im Kopieplattenlaufwerk vorhanden sind, ist das RAID-1-Niveau jedoch für die Aufrechterhaltung der Zuverlässigkeit vorteilhaft.

RAID-Niveau 2

Das RAID-Niveau 2 wird dazu verwendet, die wirtschaftlichen Nachteile des RAID-Niveaus 1 zu vermindern. Das RAID-Niveau 2 speichert Daten in verteilter Weise in jedem Plattenlaufwerk, die die Plattengruppe bilden, mit der Einheit eines Byte. Ein Hamming-Code wird dazu verwendet, einen Fehler zu erkennen und zu korrigieren. RAID-2-Niveau hat somit mehrere Prüfplattenlaufwerke zusätzlich zu den Datenplattenlaufwerken.

RAID-Niveau 3

Wenn Notwendigkeit eines Eingabe/Ausgabevorgangs ist, wird der Dateneingabe/-ausgabevorgang am Plattenlaufwerk parallel ausgeführt. Paritätsdaten werden in einem zusätzlichen Plattenlaufwerk gespeichert. Ein Spindelmotor zum Antreiben der Platte wird so synchronisiert, daß alle Plattenlaufwerke gleichzeitig Daten eingeben/ausgeben können. Es ist daher möglich, die Daten schnell zu übertragen, selbst wenn der Eingabe/Ausgabe-Vorgang der Daten nicht gleichzeitig ausgeführt wird. Wenn in einem Plattenlaufwerk ein Fehler auftritt, können die verloren gegangenen Daten durch Verwendung des Plattenlaufwerks, das normalerweise betrieben wird, und des Paritätsplattenlaufwerks wiedergewonnen werden. In diesem Falle ist die Gesamtdatenrate herabgesetzt.

Das RAID-Niveau 3 wird in einem Superrechner, einem Bildmanipulationsprozessor usw. verwendet, der eine sehr schnelle Datenübertragungsrate verlangt. Das RAID-Niveau 3 hat hohe Wirksamkeit bei der Aussendung eines langen Datenblocks (beispielsweise etwa 50 Datenblöcke), ist aber ineffektiv bei einem kurzen Datenblock (beispielsweise etwa fünf Datenblöcke).

Das RAID-Niveau 3 verwendet ein Plattenlaufwerk zur Redundanz zusammen mit dem Datenplattenlaufwerk. Das RAID-Niveau 3 benötigt daher der Zahl nach weniger Plattenlaufwerke als das RAID-Niveau 1, jedoch wird der Steuerer kompliziert und teuer.

RAID-Niveau 4

Im RAID-Niveau 4 werden Daten über mehrere Plattenlaufwerke verteilt, die eine Plattengruppe bilden. Mit anderen Worten, ein Speicherbereich jedes Plattenlaufwerks wird in mehrere Regionen unterteilt, die jeweils eine Streifengröße der Einheit eines Blocks haben, und die Daten, die der Streifengröße entsprechen, werden in jedem Plattenlaufwerk verteilt gespeichert. Paritätsdaten, die unter Verwendung der Daten berechnet werden, werden in einem zusätzlichen Plattenlaufwerk innerhalb der Plattengruppe gespeichert.

Das RAID-Niveau 4 kann wiedergewonnen werden, wenn Daten ausfallen, und seine Leseleistung ist vergleichbar dem RAID-Niveau 1. Die Schreibleistung ist jedoch beachtlich herabgesetzt im Vergleich zu einem Einzelplattenlaufwerk, weil die Paritätsinformation zu einem speziell vorgesehenen Plattenlaufwerk geleitet werden muß (in diesem Falle wird ein Flaschenhalsphänomen erzeugt). Das RAID-Niveau 4 wird durch das RAID-Niveau 5 kompensiert, dessen Schreibleistung verbessert ist.

RAID-Niveau 5

Im RAID-Niveau 5 werden die Daten in Streifen zerlegt und über alle Plattenlaufwerke der Plattengruppe verteilt. Um das Flaschenhalsphänomen beim Einschreiben zu beseitigen, werden die Paritätsdaten verteilt in allen Plattenlaufwerken gespeichert. Beim Einschreiben der Daten ist die Geschwindigkeit so niedrig wie beim RAID-Niveau 4, weil die in alle Plattenlaufwerke eingeschriebenen Daten gelesen werden müssen, um wieder die Paritätsdaten zu berechnen. Es ist jedoch möglich, die Dateneingabe/-ausgabe-Übertragung simultan auszuführen. Die Daten des ausgefallenen Plattenlaufwerks können wiedergewonnen werden.

Das RAID-5-Niveau ist wirksam beim Schreiben langer Daten. Wenn der Datenauslesung in einem Anwendungsfall mehr Gewicht gegeben wird oder wenn der Schreibleistung bei der Gestaltung der Gruppe viel Bedeutung beigemessen wird, dann kann das RAID-Niveau 5 beim Schreiben kurzer Daten wirkungsvoll sein. Wenn die Größe des Datenblocks vermindert wird, dann kann die geeignete Leistung und Datenverfügbarkeit erreicht werden. Das RAID-5-Niveau ist sehr effektiv bei den Kosten im Vergleich zu einer nichtgruppierten Vorrichtung.

Das RAID-5-Niveau hat eine Struktur ohne Datenverlust, selbst wenn ein Plattenlaufwerk, das die Plattengruppe bildet, ausfällt. Wenn das Plattenlaufwerk ausfällt und keine augenblickliche Wiedergewinnungsarbeit ausgeführt wird, dann können jedoch zusätzliche Ausfälle auftreten und somit ein Verlust an Daten hervorgerufen werden. Um den Datenverlust zu verhindern, hat das RAID-5-Niveau ein Online-Ersatzplattenlaufwerk oder ein Heißersatzplattenlaufwerk.

Fig. 5 zeigt ein Beispiel der Plattengruppe, an der die Struktur des RAID-Niveaus 5 angewendet ist. Die Plattengruppe hat fünf Plattenlaufwerke (nachfolgend als Datenlaufwerke bezeichnet) S1 bis S5 zum Speichern von Daten und ein Ersatzplattenlaufwerk (nachfolgend als Ersatzlaufwerk bezeichnet) SP. Ein Speicherbereich jedes Datenlaufwerks besteht aus n-Blöcken BLK0, ... BLKn-1. Die Größe eines Einheitsblocks wird Streifengröße genannt und hat typischerweise 512 Bytes oder ähnlich. Die Daten werden sequentiell im ersten Block BLK0 eines jeden der Datenlaufwerke S1 bis S5 gespeichert. Anschließend werden die Daten sequentiell im zweiten Block BLK1 eines jeden der Datenlaufwerke S1 bis S5 gespeichert. Die Daten werden nämlich in der Plattengruppe in der Reihenfolge vom ersten Block BLK0 des ersten Datenlaufwerks S1 über den ersten Block BLK0 des zweiten Datenlaufwerks S2 zum .... und zum ersten Block BLK0 des fünften Datenlaufwerks S5, zum zweiten Block BLK1 des ersten Datenlaufwerks S1, zum zweiten Block BLK1 des zweiten Datenlaufwerks S2, zum ... zum n-ten Block BLKn-1 des fünften Datenlaufwerks S5 gespeichert.

Wenn die Daten in den Datenlaufwerken S1 bis S5 gespeichert werden, dann werden auch Prioritätsdaten verteilt in jedem Datenlaufwerk gespeichert. In Fig. 3 sind die Paritätsdaten durch die von Kreisen umgebenen Daten im ersten Block BLK0 eines jeden Datenlauf-

werks angegeben. Das Paritätsbit ist verteilt in dem ersten Bit des ersten Datenlaufwerks S1, im zweiten Bit des zweiten Datenlaufwerks S2, dem dritten Bit des dritten Datenlaufwerks S3 und dergleichen angeordnet. Die k-ten Paritätsdaten, die verteilt in den Datenlaufwerken S1 bis S5 gespeichert sind, werden durch Exclusiv- oder -Verknüpfung der k-ten Daten von Laufwerken mit Ausnahme des Laufwerks erzeugt, in dem jene Paritätsdaten gespeichert sind. Dies wird im Detail unter Bezugnahme auf Fig. 4 erläutert.

Fig. 4 ist ein beispielhaftes Diagramm zur Beschreibung eines Verfahrens zur Erzeugung und Anordnung der Paritätsdaten in verteilter Speicherung in den Datenlaufwerken S1 bis S5. Gemäß Fig. 4 befindet sich die Paritätsdate "1" unter den ersten Bitdaten im ersten Datenlaufwerk S1 und wird durch Exclusiv- oder -Verknüpfung der ersten Bitdaten der Datenlaufwerke S2 bis S5, ausgenommen das erste Datenlaufwerk S1, erhalten. Die Paritätsdate "0" aus den zweiten Bitdaten befindet sich im zweiten Datenlaufwerk S2 und wird durch Exclusiv- oder -Verknüpfung der zweiten Bitdaten der Datenlaufwerke S1 und S3 bis S5 mit Ausnahme des zweiten Datenlaufwerks S2 erhalten. Die Paritätsdate "0" aus den dritten Bitdaten, die Paritätsdate "1" aus den vierten Bitdaten, die Paritätsdate "1" aus den fünften Bitdaten werden in der oben beschriebenen Weise erhalten. Das Verfahren zum Erzeugen der Paritätsdaten kann durch den folgenden mathematischen Ausdruck (1) dargestellt werden, in dem $\oplus$ ein Symbol ist, das eine Exclusiv- oder -Verknüpfung darstellt.

Mathematischer Ausdruck (1):

Paritätsdate des ersten Datenlaufwerks S1 = $S2 \oplus S3 \oplus S4 \oplus S5 = 1 \oplus 0 \oplus 1 \oplus 1 = 1$,

Paritätsdate des zweiten Datenlaufwerks S2 = $S1 \oplus S3 \oplus S4 \oplus S5 = 0 \oplus 1 \oplus 0 \oplus 1\ 0 = 0$,

Paritätsdate des dritten Datenlaufwerks S3 = $S1 \oplus S2 \oplus S4 \oplus S5 = 0 \oplus 1 \oplus 0 \oplus 1 = 0$,

Paritätsdate des vierten Datenlaufwerks S4 = $S1 \oplus S2 \oplus S3 \oplus S5 = 1 \oplus 1 \oplus 1 \oplus 0 = 1$, und

Paritätsdate des fünften Datenlaufwerks S5 = $S1 \oplus S2 \oplus S3 \oplus S4 = 0 \oplus 0 \oplus 0 \oplus 1 = 1$.

Die Paritätsdate, die nach dem oben beschriebenen Verfahren erzeugt wird, wird verteilt in den Datenlaufwerken S1 bis S5 gespeichert.

In der Plattengruppe mit dem Aufbau des RAID-Niveaus 5 werden die Datenlaufwerke S1 bis S5 dazu benutzt, die Daten und die Paritätsdaten zu speichern. Das Ersatzlaufwerk SP wird jedoch nicht benutzt, wenn die Plattengruppe normal betrieben wird. Das Ersatzlaufwerk SP befindet sich in einem Bereitschaftszustand während des Normalbetriebs der Plattengruppe. Wenn ein spezielles Datenlaufwerk ausfällt, wird das Ersatzlaufwerk anstelle des fehlerhaften Datenlaufwerks eingesetzt. Nimmt man an, daß das erste Datenlaufwerk S1 in Fig. 3 ausfällt, gewinnt der Plattengruppensteuerer 4 die Daten des ersten Datenlaufwerks S1 durch Exclusiv-oder -Verknüpfung der Daten der Datenlaufwerke S2 bis S5 rück und schreibt die rückgewonnenen Daten in das Ersatzplattenlaufwerk SP.

Da jedoch das Ersatzplattenlaufwerk SP innerhalb der Plattengruppe nicht verwendet wird, wenn die Plattengruppe normal betrieben wird, ist dieses unwirtschaftlich. Die Leistung des RAID-Untersystems ist folglich herabgesetzt.

Übersicht über die Erfindung

Es ist daher eine Aufgabe der vorliegenden Erfindung, ein Verfahren zum Verbessern der Fehlerausfallsicherung und der Leistung eines RAID-Untersystems anzugeben.

Eine weiteres Ziel der vorliegenden Erfindung besteht darin, ein Verfahren zum Verbessern des ineffektiven Einsatzes eines Ersatzlaufwerks in der Struktur eines RAID-Niveaus 5 anzugeben.

Eine noch weitere Aufgabe der vorliegenden Erfindung besteht darin, ein geteiltes Paritätsersatzplattenerzielungsverfahren zum Lösen eines Flaschenhalsphänomens, das in einem Aufbau des RAID-Niveaus 4 erzeugt wird, anzugeben.

Gemäß einem Aspekt der vorliegenden Erfindung wird die gestellte Aufgabe durch die im Anspruch 1 angegebenen Merkmale gelöst. Vorteilhafte Ausgestaltungen der Erfindung und weitere, dem gleichen Grundgedanken unterworfene Lösungen der Aufgabe sind Gegenstand weiterer Ansprüche.

In der vorliegenden Erfindung und Beschreibung sind ein "Datenplattenlaufwerk" zum Speichern von Daten, ein "Ersatzplattenlaufwerk", das verwendet wird, wenn ein Plattenlaufwerk ausfällt, und "Paritätsplattenlaufwerk" zur Speicherung von Paritätsdaten mit "Datenlaufwerk", "Ersatzlaufwerk" und "Paritätslaufwerk" bezeichnet.

Die vorliegende Erfindung wird unter Bezugnahme auf die Zeichnungen anhand von Ausführungsbeispielen nachfolgend näher erläutert. In den Zeichnungen sind gleiche Bezugszeichen oder Symbole für gleiche Elemente verwendet.

Kurzbeschreibung der Zeichnungen

Fig. 1 ist ein Blockdiagramm eines typischen RAID-Untersystems, das an einem Festplattenlaufwerk angewendet ist;

Fig. 2 ist ein detaillierteres Blockschaltbild eines Plattenlaufwerksteuerers nach Fig. 1;

Fig. 3 zeigt ein Beispiel einer Plattengruppe, an der die Struktur eine RAID-Niveaus 5 angewendet ist;

Fig. 4 ist ein beispielhaftes Diagramm, das der Beschreibung eines Verfahrens zum Erzeugen und zur Anordnung von Paritätsdaten dient, die verteilt in Datenlaufwerken der Struktur des RAID-Niveaus 5 gespeichert werden;

Fig. 5 ist ein Blockdiagramm eines RAID-Untersystems, das bei der vorliegenden Erfindung angewendet ist;

Fig. 6A und 6B sind ein Flußdiagramm eines Initialisierungssteuermodes, der in einem Plattengruppensteuerer gemäß der vorliegenden Erfindung ausgeführt wird;

Fig. 7 zeigt den Zustand einer Plattengruppe gemäß dem Initialisierungssteuermode der Fig. 6A und 6B;

Fig. 8 zeigt den formatierten Zustand der Plattenlaufwerke innerhalb der Plattengruppe gemäß einem Ergebnis des Initialiserungssteuermodes der Fig. 6A und 6B;

Fig. 9 ist ein Flußdiagramm, das einen Datenwiedergewinnungsvorgang zeigt, der stattfindet, wenn ein spezielles Datenlaufwerk unter den Datenlaufwerken ausfällt;

Fig. 10A und 10B sind ein Flußdiagramm, das den Datenwiedergewinnungsvorgang im Falle zeigt, daß ein Datenlaufwerk pro kleiner Paritätsgruppe ausfällt;

Fig. 11 ist ein Flußdiagramm, das einen Datenwiedergewinnungsvorgang zeigt im Falle, daß entweder ein Paritätslaufwerk oder ein Ersatzlaufwerk ausfällt;

Fig. 12 ist ein Flußdiagramm eines Datenwiedergewinnungsvorgangs im Falle, daß ein Datenlaufwerk innerhalb einer zweiten Paritätsgruppe und ein Ersatzlaufwerk ausfallen;

Fig. 13A und 13B zeigen einen Vorgang zum Lesen/Schreiben von Daten in einem RAID-Niveau 4 während einer normalen Betriebsart;

Fig. 14A und 14B zeigen einen Betrieb zum Schreiben/Lesen von Daten von/in ein Datenlaufwerk in einer ersten Paritätsgruppe während einer normalen Betriebsart gemäß der vorliegenden Erfindung; und

Fig. 15A und 15B zeigen einen Betrieb zum Schreiben von Daten in ein Datenlaufwerk in einer zweiten Paritätsgruppe während einer normalen Betriebsart gemäß der vorliegenden Erfindung.

Detaillierte Beschreibung der bevorzugten Ausführungsform

In der folgenden Beschreibung sind zahlreiche spezielle Details, wie beispielsweise die Anzahl der Plattenlaufwerke, die eine Plattengruppe, hervorgehoben, um ein vollständiges Verständnis der vorliegenden Erfindung zu bieten. Es ist für den Fachmann jedoch augenscheinlich, daß die vorliegende Erfindung auch ohne diese speziellen Details ausgeführt werden kann. Im übrigen sind gutbekannte Merkmale und Konstruktionen nicht beschrieben, weil diese für die Erläuterung der vorliegenden Erfindung nicht erforderlich sind und die Beschreibung nur belasten würden.

In der bevorzugten Ausführungsform der vorliegenden Erfindung hat eine Plattengruppe, die aus mehreren Plattenlaufwerken besteht, einen Aufbau, in dem ein RAID-Niveau 4 und ein RAID-Niveau 5 miteinander kombiniert sind. Das heißt, die Plattengruppe verwendet ein Paritätslaufwerk, das in der Struktur des RAID-Niveaus 4 verwendet wird, und ein Ersatzlaufwerk, das in der Struktur des RAID-Niveaus 5 verwendet wird, zusammen mit mehreren Datenlaufwerken.

Fig. 5 zeigt ein RAID-Untersystem, das an der vorliegenden Erfindung angewendet ist. Ein Plattengruppensteuerer 4 ist mit einem Wirts-System 2 verbunden und mit einer Plattengruppe 6 über Vielfachleitungen (BUS) verbunden. Der Plattengruppensteuerer 4 speichert in verteilter Weise Daten in jedem Plattenlaufwerk der Plattengruppe 6, führt einen Eingabe/Ausgabe-Betrieb parallel aus und wiedergewinnt bei Auftreten eines Fehlers die Daten durch Verwendung von Paritätsdaten. Eine Hauptaufgabe des Plattengruppensteuerers 4 ist es, eine Geteiltparitätsersatzplatte zu steuern. Die Steuerung der Geteiltparitätsersatzplatte, die später im Detail erläutert wird, ist in einen Initialisierungssteuermodeteil, einen normalen Steuermodeteil und einen Datenwiedergewinnungssteuermodeteil zum Restaurieren von Daten, wenn das Laufwerk ausfällt, klassifiziert.

Die Plattengruppe 6 besteht aus vier Datenlaufwerken S1 bis S4, einem Paritätslaufwerk PR und einem Ersatzlaufwerk SP. Wie in Fig. 5 gezeigt, ist die Plattengruppe 6 der Bequemlichkeit halber aus sechs Plattenlaufwerken aufgebaut. Die Konstruktion der Plattenlaufwerke der Plattengruppe 6 kann daher in Übereinstimmung mit den Wünschen des Benutzers variieren. Ein entsprechender Speicherbereich der Plattenlaufwerke S1 bis S4, PR und SP der Plattengruppe 6 ist in Blöcke mit einer Streifengröße unterteilt (beispielsweise 512 Byte).

Nachfolgend wird der Betrieb eines Initialisierungssteuermode, eines Normalbetriebsmode und eines Datenwiedergewinnungssteuermode im Detail erläutert.

Initialisierungssteuermode

Während des Initialisierungssteuermode des Plattengruppensteuerers 4 werden die Plattenlaufwerke S1 bis S4, PR und SP in kleine Paritätsgruppen unterteilt, und ein entsprechender Speicherbereich wird in einen oberen Block und einen unteren Block unterteilt. Die Fig. 6A und 6B sind ein Flußdiagramm des Initialisierungssteuermode, der in dem Plattengruppensteuerer 4 ausgeführt wird. Fig. 7 zeigt den Zustand der Plattengruppe 6 gemäß dem Initialisierungssteuermode der Fig. 6A und 6B. Fig. 8 zeigt den formatierten Zustand der Plattenlaufwerke innerhalb der Plattengruppe 6 gemäß einem Ergebnis des Initialisierungssteuermode der Fig. 6A und 6B.

Der Initialisierungssteuerbetrieb des Plattengruppensteuerers 4 und der Zustand der Plattenlaufwerke innerhalb der Plattengruppe 6 wird nun unter Bezugnahme auf die Fig. 5 bis 8 erläutert. Wenn der Plattengruppensteuerer 4 einen Systeminitialisierungsbefehl vom Wirts-System 2 empfängt, bestätigt der Plattengruppensteuerer 4 dieses im Schritt 100. Im Schritt 102 gibt der Plattengruppensteuerer 4 eine Initialisierungssteuerfunktion frei und setzt ein Teilungskennzeichen SOF. Im Schritt 104 berechnet der Plattengruppensteuerer 4 einen Zwischenzylinderwert der Plattenlaufwerke, die die Plattengruppe 6 bilden, das heißt von den Datenlaufwerken S1 bis S4, dem Paritätslaufwerk PR und dem Ersatzlaufwerk SP, und teilt jedes Plattenlaufwerk in einen oberen Block und einen unteren Block. Gemäß Fig. 7 oder 8 geben Bezugszeichen 50A, 52A, 54A, 56A, 58A und 60A die oberen Blöcke der Plattenlaufwerke an, und 50B, 52B, 54B, 56B, 58B und 60B bezeichnen die unteren Blöcke der Plattenlaufwerke. In Fig. 8 sind die oberen und unteren Blöcke der Datenlaufwerke S1 bis S4 jeweils in Blöcke unterteilt, und ihre Blockdaten UBD_0, UBD_1, ... UBD_m, LBD_0, LBD_1, ... LBD_m (wobei m ein Mehrfaches von vier ist) werden in Querrichtung verteilt. Die Blockdaten geben Daten an, die in einem Einheitsblock gespeichert sind (beispielsweise 512 Byte).

Gemäß Fig. 6A prüft der Plattengruppensteuerer 4 im Schritt 106 den Zustand des Ersatzlaufwerks SP. Im Schritt 108 prüft der Plattengruppensteuerer 4, ob ein Fehler im Ersatzlaufwerk SP vorliegt. Liegt ein Fehler vor, informiert der Plattengruppensteuerer 4 im Schritt 110 das Wirts-System 2 hierüber. Liegt kein Fehler vor, kopiert der Plattengruppensteuerer 4 im Schritt 112 die unteren Blockparitätsdaten LPR, die im unteren Block 58B des Paritätslaufwerks PR gespeichert sind, in den oberen Block 60A des Ersatzlaufwerks SP. Die im Paritätslaufwerk PR gespeicherten Paritätsdaten werden durch Exclusiv- oder -Verknüpfung der Daten der Datenlaufwerke S1 bis S4 berechnet. Gemäß Fig. 7 werden die oberen Blockdaten UBD (angedeutet durch Dreiecke) der Datenlaufwerke S1 bis S4 exclusiv- oder -verknüpft, um die oberen Blockparitätsdaten UPR zu erzeugen. Die oberen Blockparitätsdaten UPR (angedeutet durch Dreiecke) werden in dem oberen Block 58A des Paritätslaufwerks PR gespeichert. Die unteren

Blockdaten LBD (angedeutet durch Quadrate) der Datenlaufwerke S1 bis S4 werden exclusiv- oder -verknüpft, um die unteren Blockparitätsdaten LPR zu erzeugen. Die unteren Blockparitätsdaten LPR (angedeutet durch Quadrate) werden in den unteren Block 58B des Paritätslaufwerks PR gespeichert. Durch die Kopiersteuerung am Schritt 112 von Fig. 6 werden die unteren Blockparitätsdaten LPR in den oberen Block 60A des Ersatzlaufwerks SP gespeichert.

Der Plattengruppensteuerer 4 prüft am Schritt 114, ob der Kopiervorgang abgeschlossen worden ist. Wenn dieses so ist, definiert der Plattengruppensteuerer 4 am Schritt 116 die Laufwerke mit Ausnahme des Ersatzlaufwerks SP, das heißt die Datenlaufwerke S1 bis S4 und das Paritätslaufwerk PR als zwei kleine Paritätsgruppen. Beispielsweise werden die Datenlaufwerke S1 und S2 als eine erste Paritätsgruppe 30 und die Datenlaufwerke S3 und S4 und das Paritätslaufwerk PR werden als eine zweite Paritätsgruppe 40 definiert, wie in den Fig. 7 und 8 gezeigt.

Im Schritt 118 erzeugt der Plattengruppensteuerer 4 Kleingruppenoberblockparitätsdaten GUPR durch Verwendung der oberen Blockdaten UBD und UPR (in Fig. 4 durch Nullen angegeben) der Laufwerke S3, S4 und PR, die in der kleinen Paritätsgruppe einschließlich des Paritätslaufwerks PR, das heißt in der zweiten Paritätsgruppe 40, enthalten sind. Im Schritt 120 schreibt der Plattengruppensteuerer 4 die Kleingruppenoberblockparitätsdaten GUPR (in Fig. 7 durch Nullen eingezeichnet) in den unteren Block 60B des Ersatzlaufwerks SP. Der Plattengruppensteuerer 4 prüft am Schritt 122, ob der Einschreibvorgang abgeschlossen ist.

Wenn das Schreiben abgeschlossen ist, erzeugt der Plattengruppensteuerer 4 am Schritt 124 Kleingruppenunterblockparitätsdaten GLPR durch Verwendung der unteren Blockdaten LBD und LPR (in Fig. 7 durch X angedeutet) der Plattenlaufwerke S3, S4 und PR, die in der zweiten Paritätsgruppe 40 enthalten sind. Im Schritt 126 schreibt der Plattengruppensteuerer 4 die Kleingruppenunterblockparitätsdaten GLPR (in Fig. 7 durch X angedeutet) in den unteren Block 58B des Paritätslaufwerks PR. Der Plattengruppensteuerer 4 prüft im Schritt 128, ob das Einschreiben abgeschlossen worden ist. Ist dieses der Fall, setzt der Plattengruppensteuerer 4 das Teilungskennzeichen SOF rück und setzt ein Paritätsteilungsersatzkennzeichen SPSF, das anzeigt, daß das Paritätsteilen abgeschlossen worden ist, im Schritt 130. Im Schritt 132 ändert der Plattengruppensteuerer 4 den Initialisierungssteuermode zum Normalsteuermode um.

Der Zustand der Plattengruppe 6 nach dem Initialisierungssteuerbetrieb ist in den Fig. 7 und 8 abgeschlossen dargestellt. Die Oberblockparitätsdaten UPR sind im oberen Block 58A des Paritätslaufwerks PR gespeichert. Die Kleingruppenunterblockparitätsdaten GLPR sind im unteren Block 58B des Paritätslaufwerks PR gespeichert. Die unteren Blockparitätsdaten LPR sind im oberen Block 60A des Ersatzlaufwerks SP gespeichert. Die Kleingruppenunterblockparitätsdaten GUPR sind im unteren Block 60B des Ersatzlaufwerks SP gespeichert.

Nach der Beendigung des Initialisierungssteuermodes wird der Normalsteuermode ausgeführt, der nun beschrieben wird.

Normalsteuermode

Zum V rgleich, der normale Betrieb beim RAID-Niveau 4 ist unter Bezugnahme auf die Fig. 13A und 13B beschrieben. Im Datenlesebetrieb liest der Datengruppensteuerer Daten OD direkt von einem entsprechenden Datenlaufwerk S2, wie in Fig. 13A gezeigt. Ein Datenschreibvorgang ist wie folgt. Der Plattengruppensteuerer schreibt neue Daten ND in das entsprechende Datenlaufwerk S2. Zur Erzeugung der Paritätsdaten für die neuen Daten ND liest der Plattengruppensteuerer Daten OD und OP aus den Speicherbereichen aus, die dem Speicherbereich der neuen Daten ND auf den Plattenlaufwerken mit Ausnahme des Datenlaufwerks S2 entsprechen, das heißt von den Plattenlaufwerken S1, S3, S4 und PR. Die Daten OD und OP werden miteinander exclusiv- oder -verknüpft, um geänderte Daten EX zu erzeugen. Die Daten EX und die Daten ND, die neu im Datenlaufwerk S2 gespeichert sind, werden miteinander exclusiv- oder -verknüpft, um neue Paritätsdaten NP zu erzeugen. Die erzeugten Paritätsdaten NP werden im Paritätslaufwerk PR gespeichert.

Der Datenlese/Schreib-Vorgang während des Normalbetriebsmodes gemäß der vorliegenden Erfindung ist in den Fig. 14A, 14B, 15A und 15B gezeigt. In den Fig. 14A und 14B werden Daten von/in die Datenlaufwerke S2 oder S3 in der ersten Paritätsgruppe 30 während des Normalbetriebsmodes gelesen/geschrieben. In den Fig. 15A und 15B werden Daten in den Datenlaufwerken S3 oder S4 in der zweiten Paritätsgruppe 40 während des Normalbetriebsmodes gelesen/geschrieben.

Der Betrieb zum Lesen/Schreiben der Daten von/in das Datenlaufwerk S1 oder S2 in der ersten Paritätsgruppe 30 wird nun unter Bezugnahme auf die Fig. 14A und 14B erläutert. Im Datenlesebetrieb liest der Plattengruppensteuerer 4 die Daten OD direkt aus dem entsprechenden Datenlaufwerk S2 aus, wie in Fig. 14A angegeben.

Im Datenschreibbetrieb schreibt der Plattengruppensteuerer 4 die neuen Daten ND in das entsprechende Datenlaufwerk S2, wie in Fig. 14B gezeigt. Um die Paritätsdaten für die neuen Daten ND zu erzeugen, liest der Plattengruppensteuerer 4 die Daten OD aus den Speicherbereichen aus, die dem Speicherbereich der neuen Daten ND der Datenlaufwerke S1, S3 und S4 mit Ausnahme des Datenlaufwerks S2 entsprechen. Wenn der Speicherbereich der gelesenen Daten OD der obere Block ist, weil die gelesenen Daten obere Blockdaten UBD sind, dann werden die oberen Blockdaten UBD mit den oberen Blockparitätsdaten 0_UPR, die aus dem oberen Block 58A des Paritätslaufwerks PR ausgelesen werden, exclusiv- oder -verknüpft, um Daten EX1 zu erzeugen. Die Daten EX1 werden mit den neuen Daten ND, die im Datenlaufwerk S2 gespeichert sind, exclusiv-oder -verknüpft, um neue obere Blockparitätsdaten N_UPR zu erzeugen. Die erzeugten oberen Blockparitätsdaten N_UPR werden in den oberen Block 58A des Paritätslaufwerks GR eingeschrieben. Wenn der Speicherbereich der gelesenen Daten OD, die aus den Datenlaufwerken S1, S3 und S4 gelesen werden, der untere Block ist, weil die gelesenen Daten untere Blockdaten LBD sind, dann werden die unteren Blockdaten LBD mit den unteren Blockparitätsdaten 0_PR, die aus dem oberen Block 60A des Ersatzlaufwerks SP ausgelesen werden, exclusiv- oder -verknüpft, um Daten EX2 zu erzeugen. Die Daten EX2 werden mit den neuen Daten ND, die im Datenlaufwerk S2 gespeichert sind, exclusiv-oder -verknüpft, um neue unter Blockparitätsdaten N_LPR zu erzeugen. Die erzeugten unteren Blockparitätsdaten N_LPR werden in den oberen Block 60A des

Ersatzlaufwerks SP eingeschrieben.

Wenn die neuen Daten ND in eines der Datenlaufwerke S3 oder S4 eingeschrieben werden, die in der zweiten Paritätsgruppe 40 enthalten sind, denn ist der Schreibvorgang wie in den Figuren 15A und 15B gezeigt. Es wird angenommen, daß die neuen Daten ND in das Datenlaufwerk S3 eingeschrieben werden. Der Schreibvorgang von Fig. 15A ist ähnlich jenem von Fig. 14B. Die oberen und unteren Blockparitätsdaten, die in die oberen Blöcke des Paritätslaufwerks PR und des Ersatzlaufwerks SP eingeschrieben sind, werden daher in die neuen oberen und unteren Blockparitätsdaten N_UPR bzw. N_LPR geändert. Die Kleingruppenunter- und Oberblockparitätsdaten, die in die unteren Blöcke des Paritätslaufwerks PR und des Ersatzlaufwerks SP eingeschrieben sind, werden in neue Kleingruppen Unter- und -Oberblockparitätsdaten N_GLPR bzw. N_GUPR geändert. Dieser Vorgang ist in Fig. 15B dargestellt.

Bezug nehmend auf Fig. 15B: um die Paritätsdaten für die neuen Daten ND, die im Datenlaufwerk S3 gespeichert sind, zu erzeugen, werden die neuen Daten ND mit den Daten OD des Speicherbereichs, der dem Speicherbereich der neuen Daten ND im Datenlaufwerk S4 entspricht, exclusiv- oder -verknüpft, um geänderte Daten GEX zu erzeugen.

Wenn die Speicherbereiche der Daten ND und OD im oberen Block sind, weil die geänderten Daten GEX ober Blockdaten GUBD sind, werden die oberen Blockdaten GUBD mit oberen Blockparitätsdaten 0_UPR, die vom oberen Block 58A des Paritätslaufwerks PR ausgelesen werden, exclusiv- oder -verknüpft, um die neuen Kleingruppenoberblockparitätsdaten N_GUPR zu erzeugen. Die neuen Kleingruppenoberblockparitätsdaten N_GUPR werden im unteren Block 60B des Ersatzlaufwerks SP gespeichert.

Wenn die Speicherbereiche der Daten ND und OD im unteren Block sind, sind die geänderten Daten GEX untere Blockdaten GLBD. Da die unteren Blockdaten LGBD die neuen Kleingruppenunterblockparitätsdaten N_GLPR bedeuten, schreibt der Plattengruppensteuerer 4 die erzeugten Kleingruppenunterblockparitätsdaten N_GLPR in den unteren Block 58B des Paritätslaufwerks PR.

Im Falle, daß irgendein Plattenlaufwerk innerhalb der Plattengruppe 6 ausfällt, tritt ein Datenwiedergewinnungssteuermode in Aktion. Dieser wird nun im Detail erläutert.

Datenwiedergewinnungssteuermode

Während des Betriebs des RAID-Untersystems ist es möglich, Daten wiederzugewinnen im Falle, daß:

(1) ein spezifisches Datenlaufwerk unter den Datenlaufwerken S1 bis S4 ausfällt; (2) ein Datenlaufwerk pro Kleinparitätsgruppe ausfällt; (3) entweder das Paritätslaufwerk PR oder das Ersatzlaufwerk SP ausfällt; und (4) sowohl ein Datenlaufwerk innerhalb der zweiten Paritätsgruppe 40 als auch das Ersatzlaufwerk SP ausfallen.

Im Falle, daß ein spezielles Datenlaufwerk unter den Datenlaufwerken S1 bis S4 ausfällt, ist der Datenwiedergewinnungsvorgang wie in Fig. 9 gezeigt. Es sei angenommen, daß das Datenlaufwerk S3 innerhalb der Plattengruppe 6 ausfällt.

Wenn das Datenlaufwerk S1 ausfällt, ist ein Ersatzlaufwerk SP vorgesehen, um Daten des Datenlaufwerks S1 zu speichern. Die Daten des ausgefallenen Datenlaufwerks S1 werden wiedergewonnen, und die wiedergewonnenen Daten werden im Ersatzlaufwerk SP gespeichert.

Detaillierter erläutert, der Datengruppensteuerer 4 entdeckt den Ausfall am Datenlaufwerk S1 im Schritt 200 von Fig. 9. Im Schritt 202 setzt der Datengruppensteuerer 4 ein Wiedergewinnungskennzeichen RCVF. Im Schritt 204 kopiert der Datengruppensteuerer 4 die unteren Blockparitätsdaten LPR, die im oberen Block 60A des Ersatzlaufwerks SP gespeichert sind, in den unteren Block 58B des Paritätslaufwerk PR. Das Ersatzlaufwerk SP hat somit den Speicherbereich, der in der Lage ist, die Daten des Datenlaufwerks S1 zu speichern.

Der Plattengruppensteuerer 4 wiedergewinnt im Schritt 206 die Daten des ausgefallenen Datenlaufwerks S1 durch Exclusiv- oder -Verknüpfung der Daten von den Datenlaufwerken S2, S3 und S4 und des Paritätslaufwerks PR. Im Schritt 208 werden die wiedergewonnenen Daten des Datenlaufwerks S1 in das Ersatzlaufwerk SP eingeschrieben. Der Datengruppensteuerer 4 prüft im Schritt 210, ob das Einschreiben abgeschlossen worden ist. Wenn dieses der Fall ist, rekonstruiert der Plattengruppensteuerer 4 ein Laufwerkstabelle im Schritt 212. Dies bedeutet, die Laufwerkstabelle wird so rekonstruiert, daß sie das Datenlaufwerk S1 durch das Ersatzlaufwerk SP ersetzt. Im Schritt 214 wird das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt. Im Schritt 216 wird das Wiedergewinnungskennzeichen RCVF rückgesetzt. Da das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt ist, wird eine Paritätsprüfung, die das Ersatzlaufwerk SP verwendet, nicht mehr ausgeführt.

Im Falle, daß ein Datenlaufwerk pro Kleinparitätsgruppe ausfällt, wird eine Datenwiedergewinnung nach Fig. 10 ausgeführt. Es sei angenommen, daß das Datenlaufwerk S1 innerhalb der ersten Paritätsgruppe 30 und das Datenlaufwerk S3 innerhalb der zweiten Paritätsgruppe 40 ausfallen.

Wenn die Datenlaufwerke S1 und S3 ausfallen, ist ein Ersatzlaufwerk SP vorgesehen, um Daten des Datenlaufwerks S3 zu speichern. Die Daten des ausgefallenen Datenlaufwerks S3 werden unter Verwendung der Laufwerke S4 und PR, die innerhalb der zweiten Paritätsgruppe 40 nicht ausfallen, und des Ersatzlaufwerks SP wiedergewonnen. Die wiedergewonnen Daten des ausgefallenen Datenlaufwerks S3 werden im Ersatzlaufwerk SP gespeichert. Die Daten des ausgefallenen Datenlaufwerks S1 werden unter Verwendung der Datenlaufwerke S2 und S4, des Paritätslaufwerks PR und des Ersatzlaufwerks SP, das die wiedergewonnenen Daten des Datenlaufwerks S3 speichert, wiedergewonnen. Die wiedergewonnenen Daten des ausgefallenen Datenlaufwerks S1 werden im Paritätslaufwerk PR gespeichert.

Detaillierter erläutert, der Plattengruppensteuerer 4 ermittelt den Ausfall der Datenlaufwerke S1 und S3 aus jeder Paritätsgruppe im Schritt 300 von Fig. 10. Im Schritt 302 setzt der Plattengruppensteuerer 4 das Wiedergewinnungskennzeichen RCVF. Im Schritt 304 setzt der Plattengruppensteuerer 4 ein Ersatzparitätskennzeichen RPPF. Im Schritt 306 wechselt der Plattengruppensteuerer 4 die unteren Blockparitätsdaten LPR, die im oberen Block 60a des Ersatzlaufwerks SP gespeichert sind, und die Kleingruppenunterblockparitätsdaten GLPR, die im unteren Block 58B des Paritätslaufwerks PR gespeichert sind, gegeneinander aus. Die unteren Blockparitätsdaten LPR werden daher im unteren

Block 58B des Paritätslaufwerks PR gespeichert, und die Kleingruppenunterblockparitätsdaten GLPR werden im oberen Block 60A des Ersatzlaufwerks SP gespeichert. Im Schritt 308 setzt der Plattengruppensteuerer 4 das Ersatzparitätskennzeichen RPPF rück.

Im Schritt 310 wiedergewinnt der Plattengruppensteuerer 4 die Daten des ausgefallenen Datenlaufwerks S3 innerhalb der zweiten Paritätsgruppe 40 durch Verwendung der Daten vom Datenlaufwerk S4, dem Paritätslaufwerk PR und dem Ersatzlaufwerk SP. Die oberen Blockdaten des ausgefallenen Datenlaufwerks S3 werden nämlich wiedergewonnen durch Exclusiv- oder -Verknüpfung der oberen Blockdaten UBD des Datenlaufwerks S4, der oberen Blockparitätsdaten UPR des Paritätslaufwerks PR und der Kleingruppenoberblockparitätsdaten GUPR, die im unteren Block 60B des Ersatzlaufwerks SP gespeichert sind. Die unteren Blockdaten des ausgefallenen Datenlaufwerks S3 werden wiedergewonnen durch Exclusiv- oder -Verknüpfung der unteren Blockdaten LBD des Datenlaufwerks S4, der unteren Blockparitätsdaten LPR des Paritätslaufwerks PR und der Kleingruppenunterblockparitätsdaten GLPR, die im oberen Block 60A des Ersatzlaufwerks SP gespeichert sind. Im Schritt 312 schreibt der Plattengruppensteuerer 4 die wiedergewonnenen Daten, das heißt die oberen und unteren Blockdaten des ausgefallenen Datenlaufwerks S3 in das Ersatzlaufwerk SP ein. Der Plattengruppensteuerer 4 prüft im Schritt 314, ob das Einschreiben abgeschlossen worden ist.

Ist das Einschreiben abgeschlossen worden, wiedergewinnt der Plattengruppensteuerer 4 im Schritt 316 die Daten des ausgefallenen Datenlaufwerks S1 innerhalb der ersten Paritätsgruppe 30 durch Verwendung der Daten des Paritätslaufwerks PR, der Datenlaufwerke S2 und S4 und des Ersatzlaufwerks SP, das die wiedergewonnenen Daten des ausgefallenen Datenlaufwerks S3 speichert. Das heißt, die oberen Blockdaten des ausgefallenen Datenlaufwerks S1 werden wiedergewonnen durch Exclusiv- oder -Verknüpfung der oberen Blockdaten UBD der Datenlaufwerke S1 und S4, der oberen Blockparitätsdaten UPR des Paritätslaufwerks PR und der wiedergewonnenen oberen Blockdaten des Datenlaufwerks S3, die im oberen Block 60A des Ersatzlaufwerks SP gespeichert sind. Die unseren Blockdaten des ausgefallenen Datenlaufwerks S1 werden durch Exclusiv- oder -Verknüpfung der unteren Blockdaten LBD der Datenlaufwerke S1 und S3, der unteren Blockparitätsdaten LPR des Paritätslaufwerks PR und der wiedergewonnenen unteren Blockdaten des Datenlaufwerks S3, die im unteren Block 30B des Ersatzlaufwerks SP gespeichert sind, wiedergewonnen.

Der Plattengruppensteuerer 4 schreibt die wiedergewonnenen oberen und unteren Blockdaten des ausgefallenen Datenlaufwerks S1 in das Paritätslaufwerk PR im Schritt 316. Im Schritt 320 wird dann geprüft, ob das Einschreiben abgeschlossen worden ist. Wenn letzteres der Fall ist, rekonstruiert der Plattengruppensteuerer 4 die Laufwerkstabelle. Die Laufwerkstabelle wird nämlich rekonstruiert, um das Datenlaufwerk S1 durch das Paritätslaufwerk PR und das Datenlaufwerks S3 durch das Ersatzlaufwerk SP zu ersetzen. Im Schritt 324 setzt der Plattengruppensteuerer 4 das Geteiltparitätsersatzkennzeichen SPSF rück. Im Schritt 326 setzt der Plattengruppensteuerer 4 das Wiedergewinnungskennzeichen RCVF rück. Da das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt ist, wird die Paritätsprüfung unter Verwendung des Paritätslaufwerks PR und des Ersatzlaufwerks SP nicht ausgeführt.

Im Falle, daß entweder das Paritätslaufwerk oder das Ersatzlaufwerk SP ausfällt, ist der Datenwiedergewinnungsvorgang wie in **Fig.** 11 gezeigt.

Wenn das Paritätslaufwerk PR ausfällt, werden die oberen Blockparitätsdaten UPR des Paritätslaufwerks PR durch Verwendung der oberen Blockdaten UBD der Datenlaufwerke S1, S2, S3 und S4 rückgewonnen, weil die unteren Blockparitätsdaten LPF im oberen Block 60A des Ersatzlaufwerks SP gespeichert sind. Die wiedergewonnenen Daten werden im Ersatzlaufwerk SP gespeichert. Wenn das Ersatzlaufwerk SP ausfällt, werden die unteren Blockparitätsdaten LPR durch Verwendung der unteren Blockdaten LBD der Datenlaufwerke S1, S2, S3 und S4 wiedergewonnen, weil nur die unteren Blockparitätsdaten LPR, die im Ersatzlaufwerk SP gespeichert sind, ausfallen. Die wiedergewonnenen Daten werden im Paritätslaufwerk PR gespeichert.

Detaillierter erläutert, der Plattengruppensteuerer 4 ermittelt das ausgefallene Laufwerk (das heißt das Paritätslaufwerk PR oder das Ersatzlaufwerk SP) im Schritt 400, wie in **Fig.** 11 gezeigt. Im Schritt 402 setzt der Plattengruppensteuerer 4 das Rückgewinnungskennzeichen RCVF. Der Plattengruppensteuerer 4 prüft im Schritt 404, ob das Paritätslaufwerk PR ausfällt. Ist dieses der Fall, kopiert der Plattengruppensteuerer 4 im Schritt 406 die unteren Blockparitätsdaten PR, die im oberen Block 60A des Ersatzlaufwerks SP gespeichert sind, in dessen unteren Block 60B. Im Schritt 408 wiedergewinnt der Plattengruppensteuerer 4 die oberen Blockparitätsdaten UPR des Paritätslaufwerks PR durch Exclusiv-oder -Verknüpfung der oberen Blockdaten UBD aller Datenlaufwerke S1, S2, S3 und S4 innerhalb der Plattengruppe 6. Die wiedergewonnenen oberen Blockparitätsdaten UPR werden in den oberen Block 60A des Ersatzlaufwerks SP im Schritt 410 eingeschrieben. Im Schritt 412 wird geprüft, ob das Einschreiben abgeschlossen worden ist. Wenn dieses der Fall ist, rekonstruiert der Plattengruppensteuerer 4 im Schritt 414 die Laufwerkstabelle, um das Paritätslaufwerk PR durch das Ersatzlaufwerk SP zu ersetzen.

Wenn indessen das Ersatzlaufwerk SP im Schritt 404 ausfällt, wiedergewinnt der Plattengruppensteuerer 4 im Schritt 416 die unteren Blockparitätsdaten LPR durch Exclusiv- oder -Verknüpfung der unteren Blockdaten LBD aller Datenlaufwerke S1, S2, S3 und S4 innerhalb der Plattengruppe 6. Die wiedergewonnenen unteren Blockparitätsdaten LPR werden in den unteren Block 58B des Paritätslaufwerks PR im Schritt 418 eingeschrieben. Im Schritt 420 wird geprüft, ob der Einschreibvorgang abgeschlossen ist. Wenn dieses der Fall ist, rekonstruiert der Plattengruppensteuerer 4 im Schritt 422 die Laufwerkstabelle so, daß das Ersatzlaufwerk SP nicht verwendet wird. Im Schritt 424 wird das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt, und im Schritt 426 wird das Rückgewinnungskennzeichen RCVF rückgesetzt.

Im Falle, daß ein Plattenlaufwerk innerhalb der zweiten Paritätsgruppe 40 und das Ersatzlaufwerk SP ausfallen, erfolgt die Datenrückgewinnung nach **Fig.** 12. Es sei angenommen, daß das Datenlaufwerk S3 innerhalb der zweiten Gruppe 40 und das Ersatzlaufwerk SP ausfallen.

Wenn das Datenlaufwerk S3 und das Ersatzlaufwerk SP ausfallen, werden die oberen Blockdaten des ausgefallenen Datenlaufwerks S3 unter Verwendung der oberen Blockdaten UBD und UPR der Datenlaufwerke S1, S2 und S3 und des Paritätslaufwerks PR wiedergewonnen. Die wiedergewonnenen oberen Blockdaten werden

im oberen Block 58A des Paritätslaufwerks PR gespeichert. Die unteren Blockdaten das ausgefallenen Datenlaufwerks S3 werden unter Verwendung der unteren Blockdaten LBD des Datenlaufwerks S4 und der Kleingruppenunterblockparitätsdaten GLPR, die im unteren Block 58B des Paritätslaufwerks PR gespeichert sind, wiedergewonnen. Die wiedergewonnenen unteren Blockdaten werden im unteren Block 58B des Paritätslaufwerks PR gespeichert.

Detaillierter erläutert, der Plattengruppensteuerer 4 ermittelt das ausgefallene Datenlaufwerk S3 und das ausgefallene Paritätslaufwerk PR im Schritt 500 von Fig. 12. Im Schritt 502 setzt der Plattengruppensteuerer das Wiedergewinnungskennzeichen RCVF.

Im Schritt 504 wiedergewinnt der Plattengruppensteuerer 4 die oberen Blockdaten des ausgefallenen Datenlaufwerks S3 durch Exclusiv- oder -Verknüpfung von UBD und UPR der Datenlaufwerke S1, S2 und S4 und des Paritätslaufwerks PR, die innerhalb der Plattengruppe 6 nicht ausfallen. Im Schritt 506 werden die wiedergewonnenen oberen Blockdaten des Datenlaufwerks S3 in den oberen Block 58A des Paritätslaufwerks PR eingeschrieben. Im Schritt 508 wird geprüft, ob das Einschreiben abgeschlossen ist.

Wenn dieses der Fall ist, wiedergewinnt der Plattengruppensteuerer 4 im Schritt 510 die unteren Blockdaten des ausgefallenen Datenlaufwerks S3 durch Exclusiv- oder -Verknüpfung der unteren Blockdaten LBD des Datenlaufwerks S4, das innerhalb der zweiten Paritätsgruppe 40 nicht ausfällt, mit den Kleingruppenunterparitätsdaten GLPR, die im unteren Block 58B des Paritätslaufwerks PR gespeichert sind. Die wiedergewonnenen unteren Blockdaten des Datenlaufwerks S3 werden in den unteren Block 58B des Paritätslaufwerks im Schritt 512 eingeschrieben. Der Plattengruppensteuerer 4 prüft, ob das Einschreiben abgeschlossen worden ist, und zwar im Schritt 514. Ist dieses der Fall, rekonstruiert der Plattengruppensteuerer 4 die Laufwerkstabelle, um das Datenlaufwerk S3 durch das Paritätslaufwerk PR zu ersetzen. Im Schritt 518 wird das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt, und im Schritt 520 wird das Wiedergewinnungskennzeichen RCVF rückgesetzt. Da das Geteiltparitätsersatzkennzeichen SPSF rückgesetzt ist, wird eine Paritätsprüfung, die das Paritätslaufwerk PR und das Ersatzlaufwerk SP verwendet, nicht ausgeführt.

Wie oben erwähnt, hat das erfindungsgemäße Verfahren zur Erzielung einer geteilten Paritäts/Ersatz-Platte die folgenden Vorteile: zunächst werden die Quellen wirksam verwaltet, weil die Ersatzplatte auch im Normalbetrieb verwendet wird. Zweitens ist die Auslastung des Paritätslaufwerks herabgesetzt, und die Gesamtleistung des Systems ist verbessert, indem die Paritätsdaten des Paritätslaufwerks verteilt gespeichert werden. Drittens ist es möglich, die Daten unter vorhersehbaren Ausfallumständen wiederzugewinnen, indem das Datenlaufwerk auf Kleinparitätsgruppen aufgeteilt wird und der Paritätsresultatswert in einem unbenutzten Bereich gespeichert wird.

Es versteht sich, daß die vorliegende Erfindung nicht auf die hier speziell als beste erläuterte Ausführungsform beschränkt ist, sondern daß sie auch davon abweichend realisiert werden kann, solange sie unter dem Geist der anhängenden Ansprüche fällt.

Patentansprüche

1. Verfahren zum Verbessern der Fehlerfestigkeit und Leistungsfähigkeit eines RAID-Untersystems, das Daten in einer Plattengruppe, die aus mehreren Plattenlaufwerken besteht, verteilt speichert und einen parallelen Eingabe/Ausgabebetrieb ausführt, umfassend die folgenden Schritte:
Aufbauen der Plattengruppe mit wenigstens zwei Datenplattenlaufwerken zum Speichern von Daten, einem Ersatzplattenlaufwerk, das beim Ausfall eines Plattenlaufwerks eingesetzt wird, und ein Paritätsplattenlaufwerk zum Speichern von Paritätsdaten; und
das Aufteilen der Paritätsdaten des Paritätsplattenlaufwerks und Speichern der geteilten Daten in dem Paritätsplattenlaufwerk und dem Ersatzplattenlaufwerk.

2. Verfahren nach Anspruch 1, bei dem die Plattengruppe 4 Datenplattenlaufwerke, ein Paritätsplattenlaufwerk und ein Ersatzplattenlaufwerk umfaßt.

3. Verfahren zur Verbesserung der Fehlerfestigkeit und Leistungsfähigkeit eines RAID-Untersystems, das Daten in einer Plattengruppe, die aus einem Paritätsplattenlaufwerk, einem Ersatzplattenlaufwerk und mehreren Datenplattenlaufwerken besteht, verteilt speichert und einen parallelen Eingabe/Ausgabebetrieb ausführt, umfassend die folgenden Schritte:
Teilen von Paritätsdaten, die in dem Paritätsplattenlaufwerk gespeichert sind, in obere Blockparitätsdaten und untere Blockparitätsdaten und Speichern der oberen Blockparitätsdaten und der unteren Blockparitätsdaten in einen oberen Blockspeicherbereich des Paritätsplattenlaufwerks und einen unteren Blockspeicherbereich des Ersatzplattenlaufwerks,
Definieren der vorbeschriebenen Anzahl von Plattenlaufwerken mit Ausnahme des Ersatzplattenlaufwerks als Kleinparitätsgruppen und Erzeugen von Kleingruppenparitätsdaten unter Verwendung von Plattenlaufwerken einer Kleinparitätsgruppe, die das Paritätsplattenlaufwerk enthält, und
Teilen der Kleingruppenparitätsdaten in Kleingruppenoberblockparitätsdaten und Kleingruppenunterblockparitätsdaten und Speichern der Kleingruppenoberblockparitätsdaten und der Kleingruppenunterblockparitätsdaten in einen unteren Blockspeicherbereich des Ersatzplattenlaufwerks bzw. einen unteren Blockspeicherbereich des Paritätsplattenlaufwerks.

4. Verfahren nach Anspruch 3, bei dem die Kleinparitätsgruppe enthält:
eine erste Paritätsgruppe, die aus mehreren Datenplattenlaufwerken besteht und das Paritätsplattenlaufwerk nicht enthält, und
eine zweite Paritätsgruppe, die aus mehreren Datenplattenlaufwerken und dem Paritätsplattenlaufwerk besteht.

5. Verfahren nach Anspruch 4, bei dem die Plattenlaufwerke in einen oberen Block und einen unteren Block auf der Basis eines Zwischenzylinderwertes jedes Plattenlaufwerks unterteilt sind.

6. Datenwiedergewinnungsverfahren für den Fall des Ausfalls eines Datenplattenlaufwerks, das nicht in einer Kleinparitätsgruppe enthalten ist, in einem RAID-Untersystem, das eine Plattengruppe enthält, die besteht aus: mehreren Datenplattenlaufwerken zum Speichern von Daten in oberen und unteren Blockspeicherbereichen; einem Paritätsplattenlaufwerk zum Speichern in seinem oberen

Blockspeicherbereich obere Blockparitätsdaten für obere Blockdaten, die in oberen Blockspeicherbereichen der Datenplattenlaufwerke gespeichert sind, und zum Speichern in ihrem unteren Blockspeicherbereich Kleingruppenunterblockparitätsdaten für untere Blockdaten, die in unteren Blockspeicherbereichen von Datenplattenlaufwerken innerhalb der Kleinparitätsgruppe gespeichert sind, die so definiert sind, daß Kleingruppenober- und -unterblockparitätsdaten erzeugt werden; und ein Ersatzplattenlaufwerk zum Speichern in seinem oberen Blockspeicherbereich untere Blockparitätsdaten für untere Blockdaten, die in unteren Blockspeicherbereichen der Datenplattenlaufwerke gespeichert sind und zum Speichern in ihrem unteren Blockspeicherbereich von Kleingruppenoberparitätsdaten für obere Blockdaten, die in oberen Blockspeicherbereichen der Datenplattenlaufwerke innerhalb der Kleinparitätsgruppe und des Paritätsplattenlaufwerks gespeichert sind, wobei das Verfahren die Schritte umfaßt:

Kopieren der unteren Blockparitätsdaten, die in dem oberen Blockspeicherbereich des Ersatzlaufwerks gespeichert sind, in den unteren Blockspeicherbereich des Paritätslaufwerks, um obere und untere Blockdaten des ausgefallenen Datenplattenlaufwerks zu speichern;

Wiedergewinnen der oberen und unteren Blockdaten des ausgefallenen Datenplattenlaufwerks durch Verwendung der oberen und unteren Blockdaten anderer Plattenlaufwerke der Plattengruppe, die nicht ausfallen, und

Einschreiben der wiedergewonnenen oberen und unteren Blockdaten in die oberen und unteren Blockspeicherbereiche des Ersatzplattenlaufwerks.

7. Datenwiedergewinnungsverfahren für den Fall, daß ein Datenplattenlaufwerk innerhalb einer kleinen Paritätsgruppe und ein Datenplattenlaufwerk, das nicht in der Kleinparitätsgruppe enthalten ist, in einem RAID-Untersystem ausfallen, wobei die Plattengruppe enthält: mehrere Datenplattenlaufwerke zum Speichern von Daten in oberen und unteren Blockspeicherbereichen, ein Paritätsplattenlaufwerk zum Speichern in seinem oberen Blockspeicherbereich von oberen Blockparitätsdaten für obere Blockdaten, die in oberen Blockspeicherbereichen der Datenplattenlaufwerke gespeichert sind, und zum Speichern in seinem unteren Blockspeicherbereich von Kleingruppenunterblockparitätsdaten für untere Blockdaten, die in unteren Blockspeicherbereichen von Datenplattenlaufwerken innerhalb der Kleinparitätsgruppe gespeichert sind, die definiert sind, um Kleingruppenober- und -unterblockparitätsdaten zu erzeugen; und ein Ersatzplattenlaufwerk zum Speichern in seinem oberen Blockspeicherbereich von unteren Blockparitätsdaten für untere Blockdaten, die in unteren Blockspeicherbereichen der Datenplattenlaufwerke gespeichert sind und zum Speichern in seinem oberen Speicherbereich von Kleingruppenoberblockparitätsdaten über obere Blockdaten, die in oberen Blockspeicherbereichen von Datenplattenlaufwerken innerhalb der Kleinparitätsgruppe und des Paritätsplattenlaufwerks gespeichert sind, wobei das Verfahren die folgenden Schritte umfaßt:

Austauschen der unteren Blockparitätsdaten, die in dem oberen Blockspeicherbereich des Ersatzplattenlaufwerks gespeichert sind, und der Kleingruppenunterblockparitätsdaten, die in dem unteren Blockspeicherbereich des Paritätsplattenlaufwerks gespeichert sind, gegeneinander;

Rückgewinnen der oberen und unteren Blockdaten des ausgefallenen Datenplattenlaufwerks innerhalb der Kleinparitätsgruppe durch Verwendung der oberen und unteren Blockdaten des Datenplattenlaufwerks, das innerhalb der Kleinparitätsgruppe, des Paritätsplattenlaufwerks und des Ersatzplattenlaufwerks nicht ausfällt;

Einschreiben der wiedergewonnenen oberen und unteren Blockdaten in die oberen und unteren Blockspeicherbereiche des Ersatzplattenlaufwerks;

Wiedergewinnen der oberen und unteren Blockdaten des ausgefallenen Datenplattenlaufwerks, das nicht in der Kleinparitätsgruppe enthalten ist, durch Verwenden der oberen und unteren Blockdaten der nicht ausfallenden Datenplattenlaufwerke des Paritätsplattenlaufwerks und des Ersatzplattenlaufwerks, und

Einschreiben der wiedergewonnenen oberen und unteren Blockdaten des ausgefallenen Datenplattenlaufwerks, das nicht in der Kleinparitätsgruppe enthalten ist, in die oberen und unteren Blockspeicherbereiche des Paritätsplattenlaufwerks.

Hierzu 17 Seite(n) Zeichnungen

– Leerseite –

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

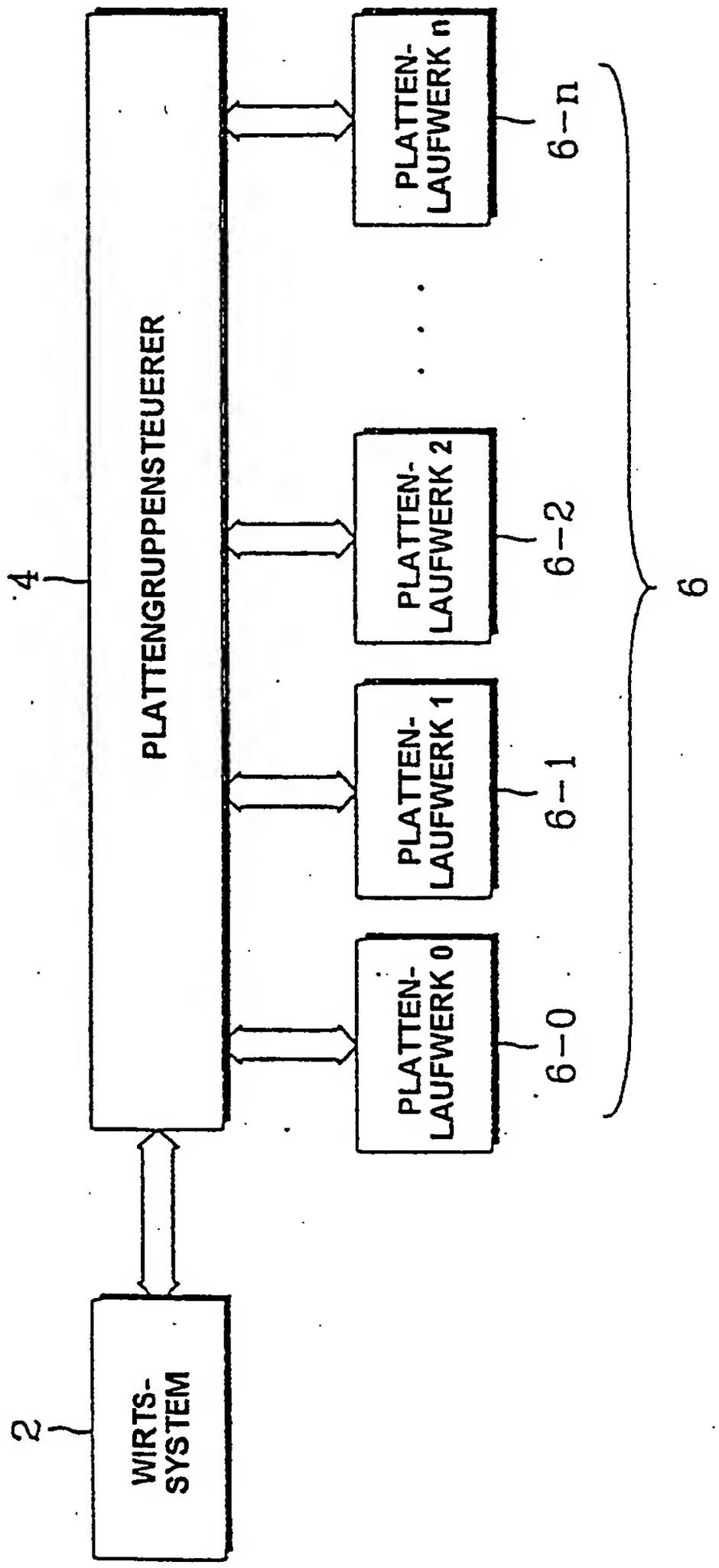


Fig. 1

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

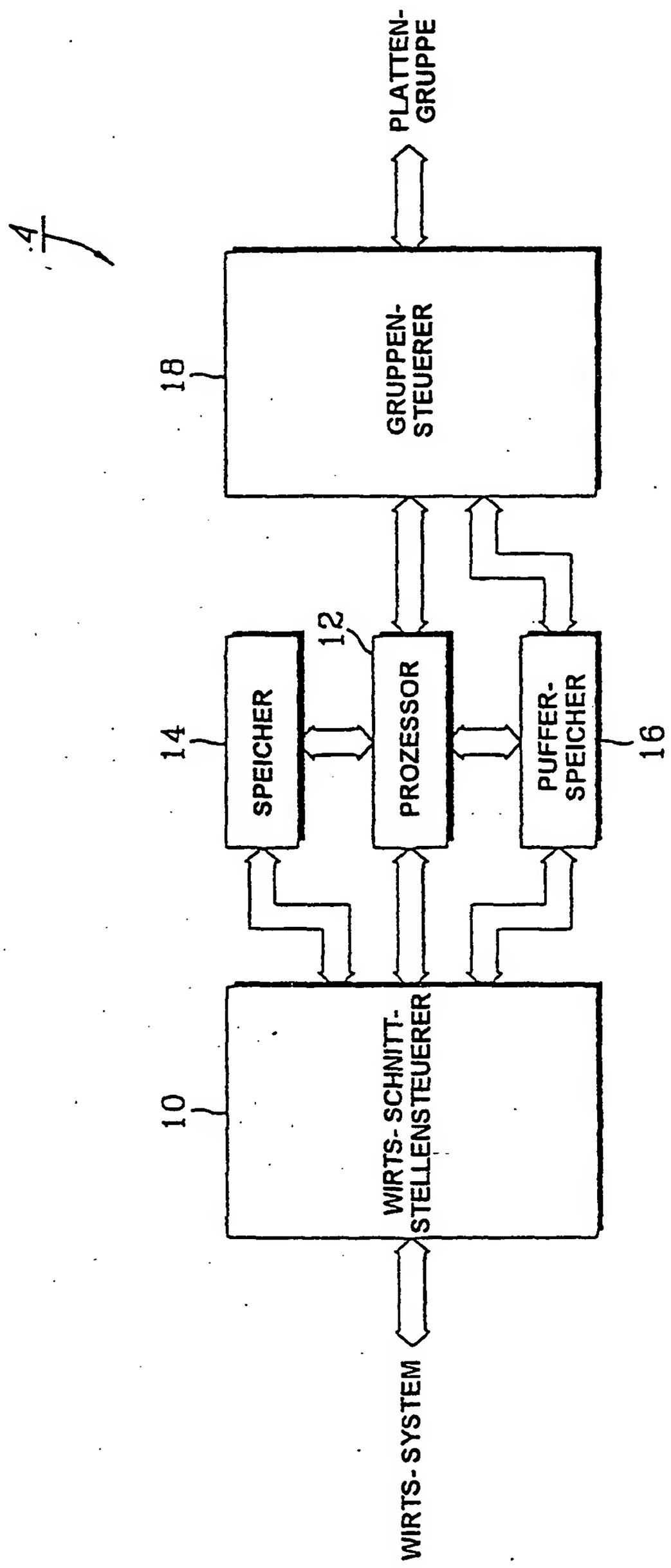


Fig. 2

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

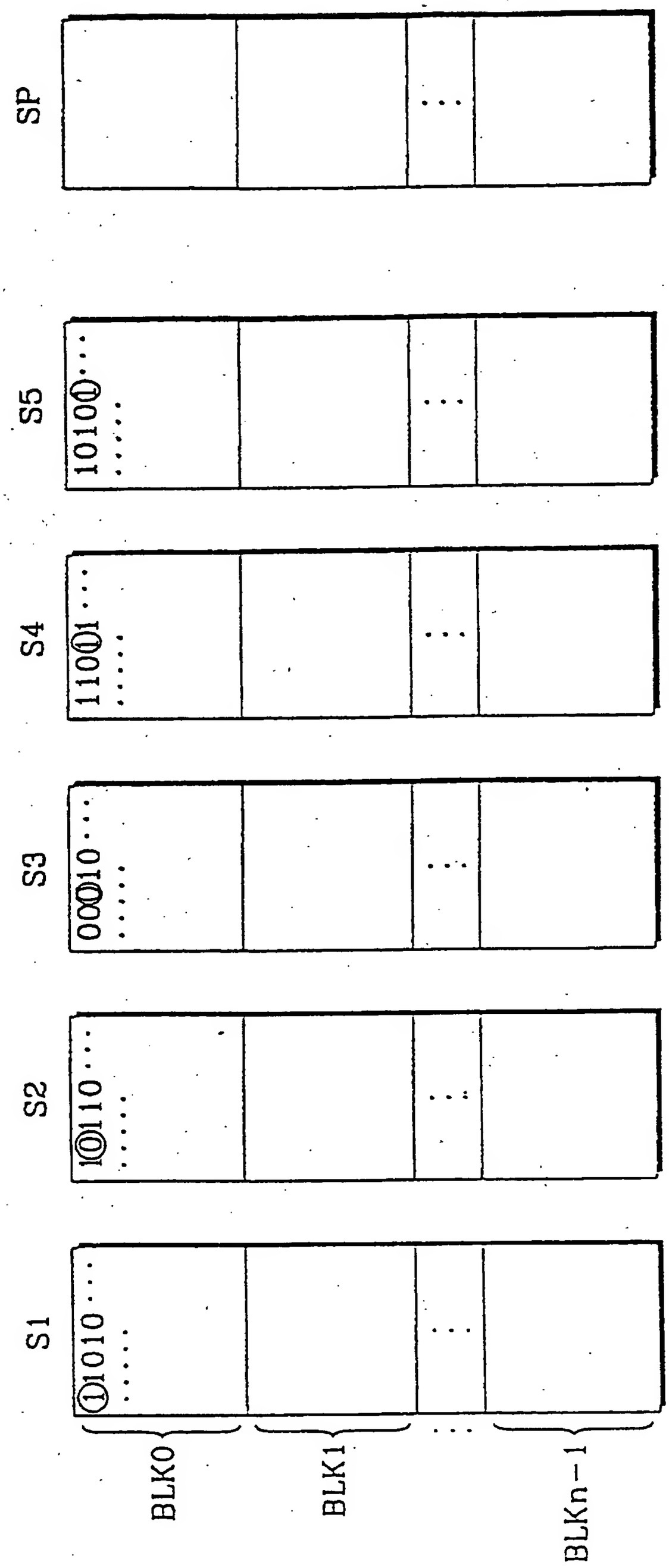


Fig. 3

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

	S1	S2	S3	S4	S5	SP
ERSTE BITDATEN	①	1	0	1	1	
ZWEITE BITDATEN	1	⓪	0	1	0	
DRITTE BITDATEN	0	1	⓪	0	1	
VIERTE BITDATEN	1	1	1	①	0	
FÜNFTE BITDATEN	0	0	0	1	①	

Fig. 4

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

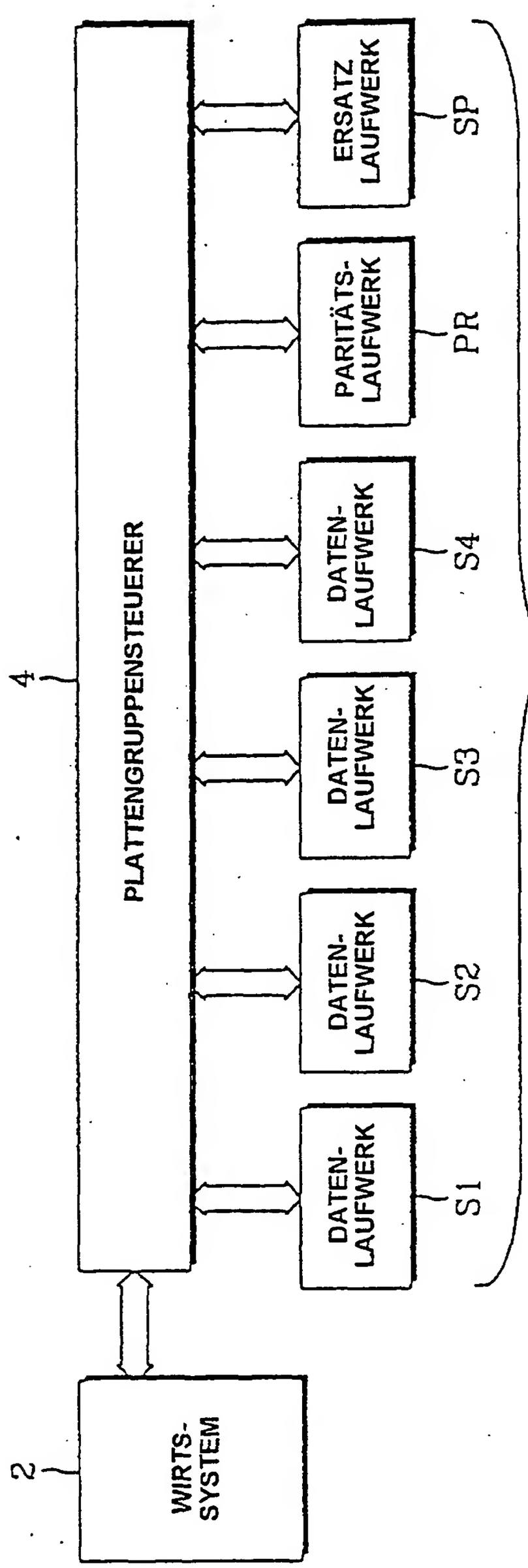


Fig. 5

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

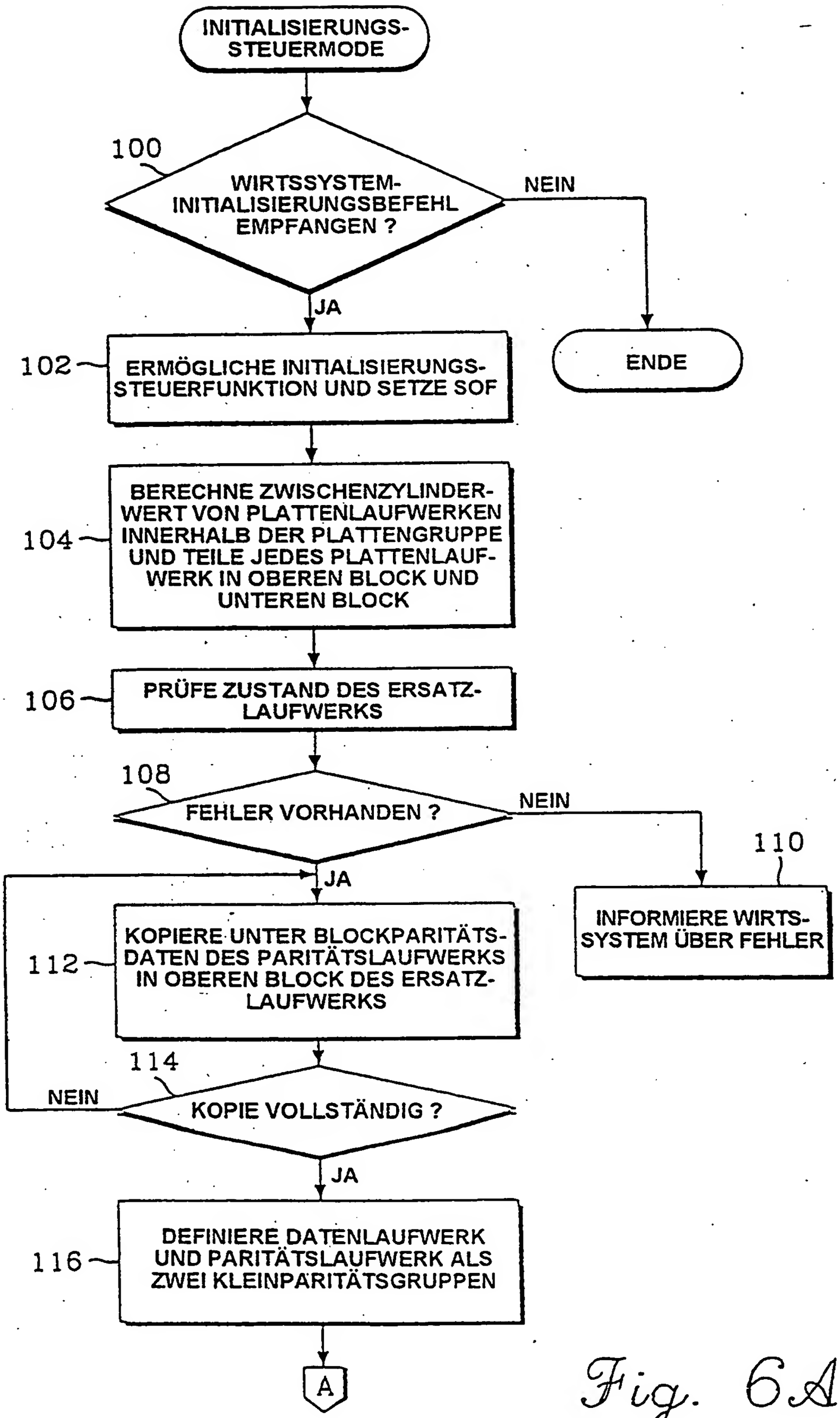


Fig. 6A

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

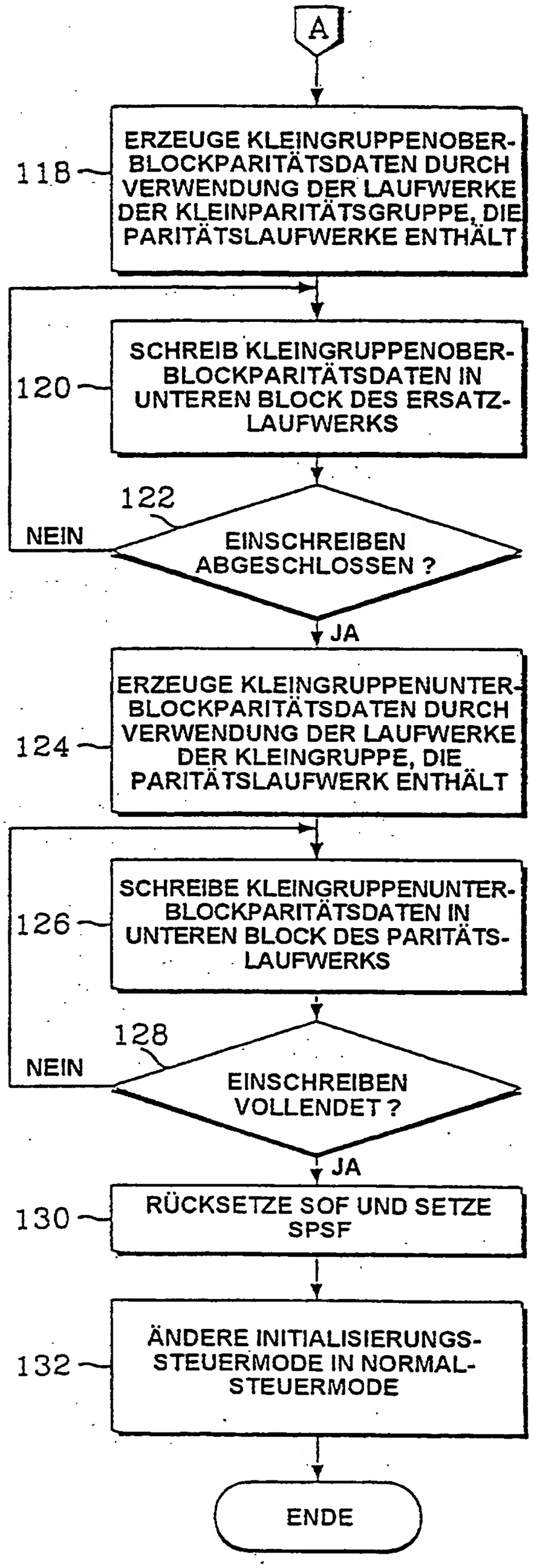


Fig. 6B

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

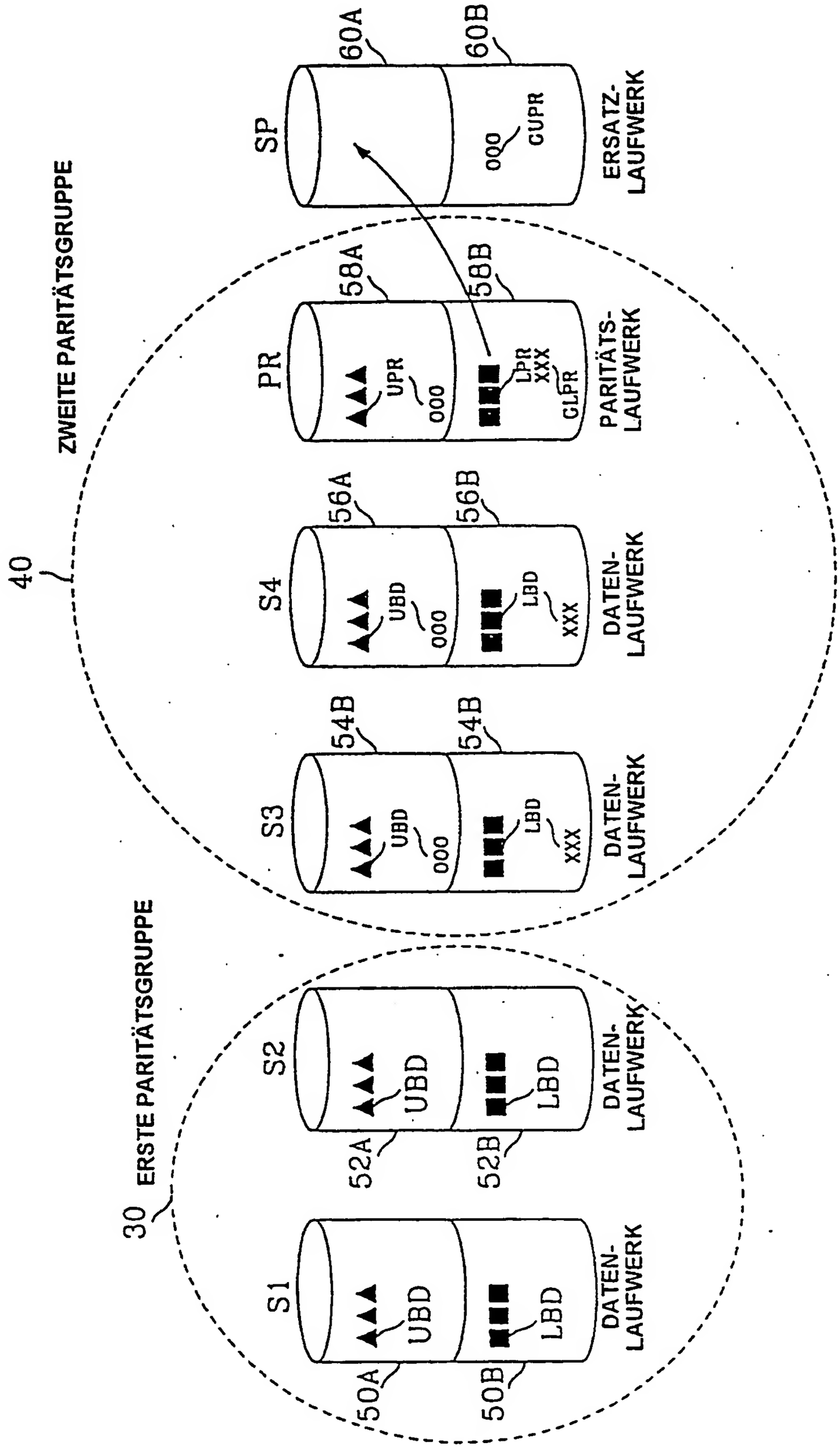


Fig. 7

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

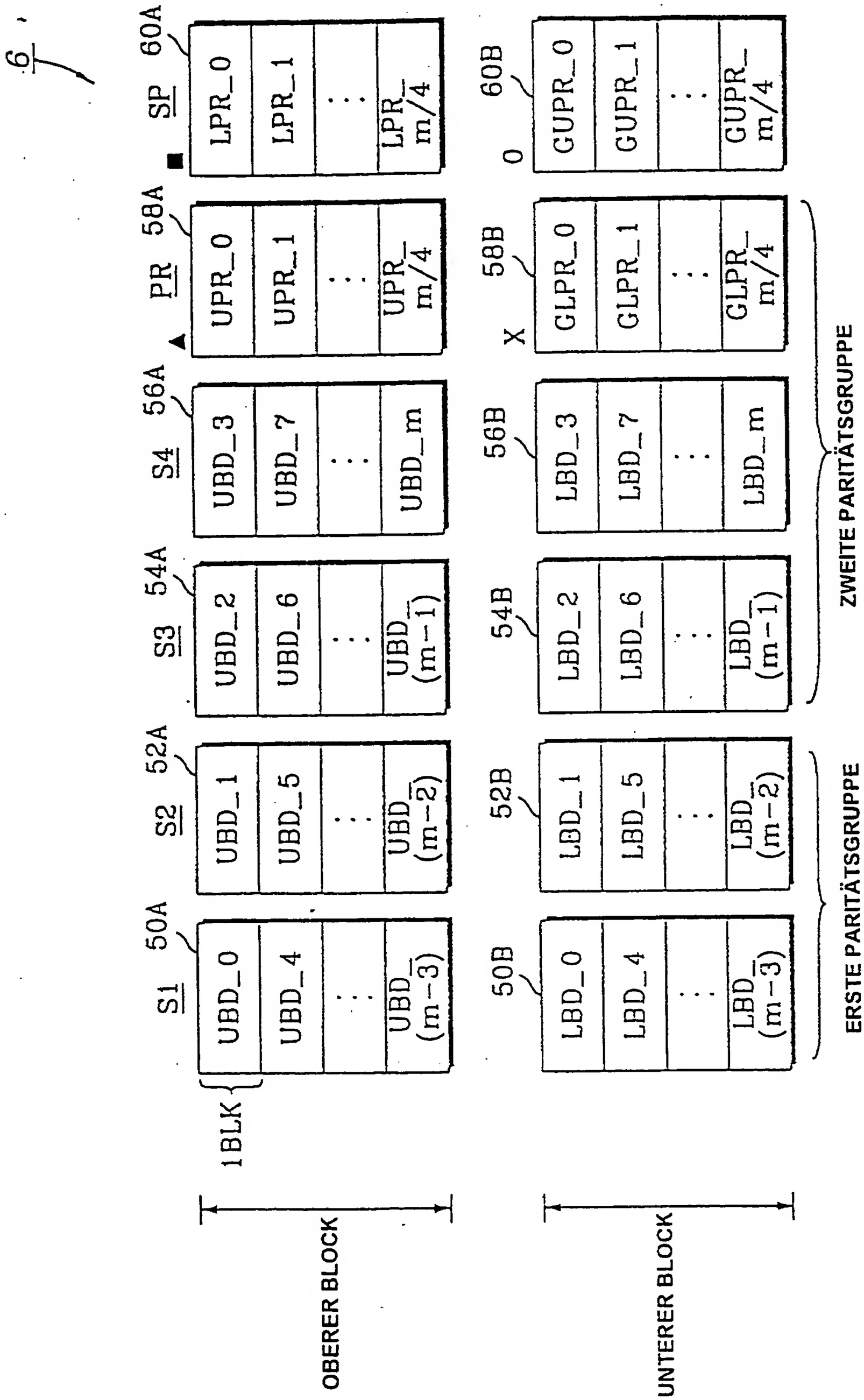


Fig. 8

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

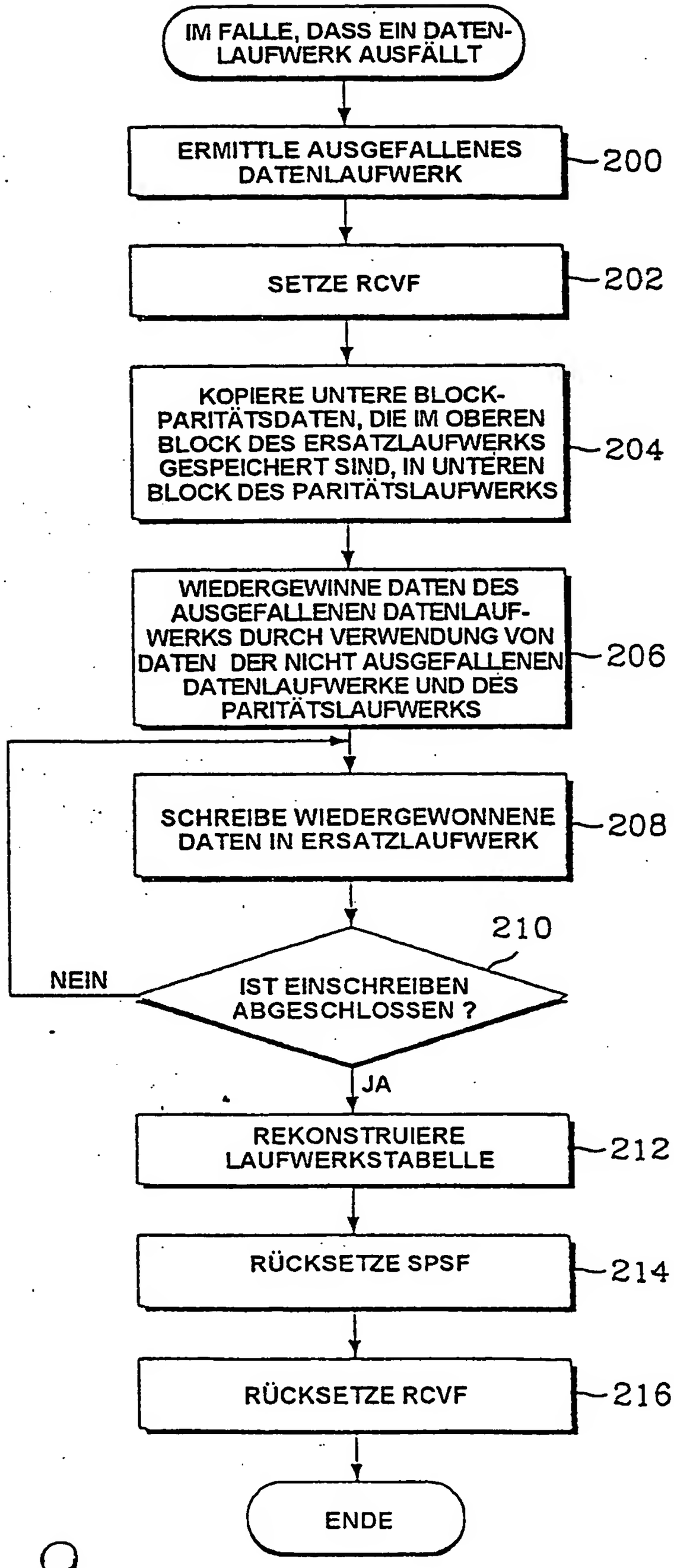


Fig. 9

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

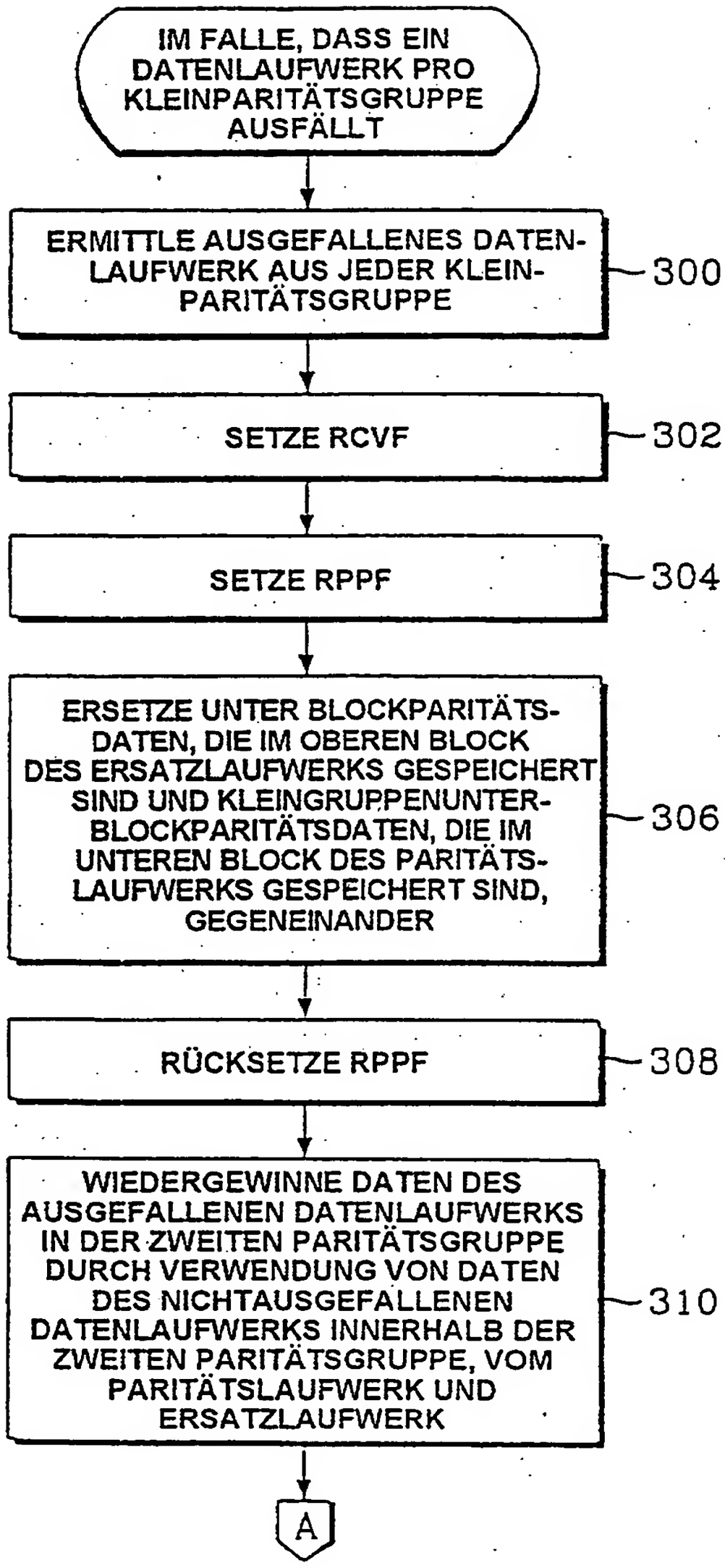


Fig. 10A

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

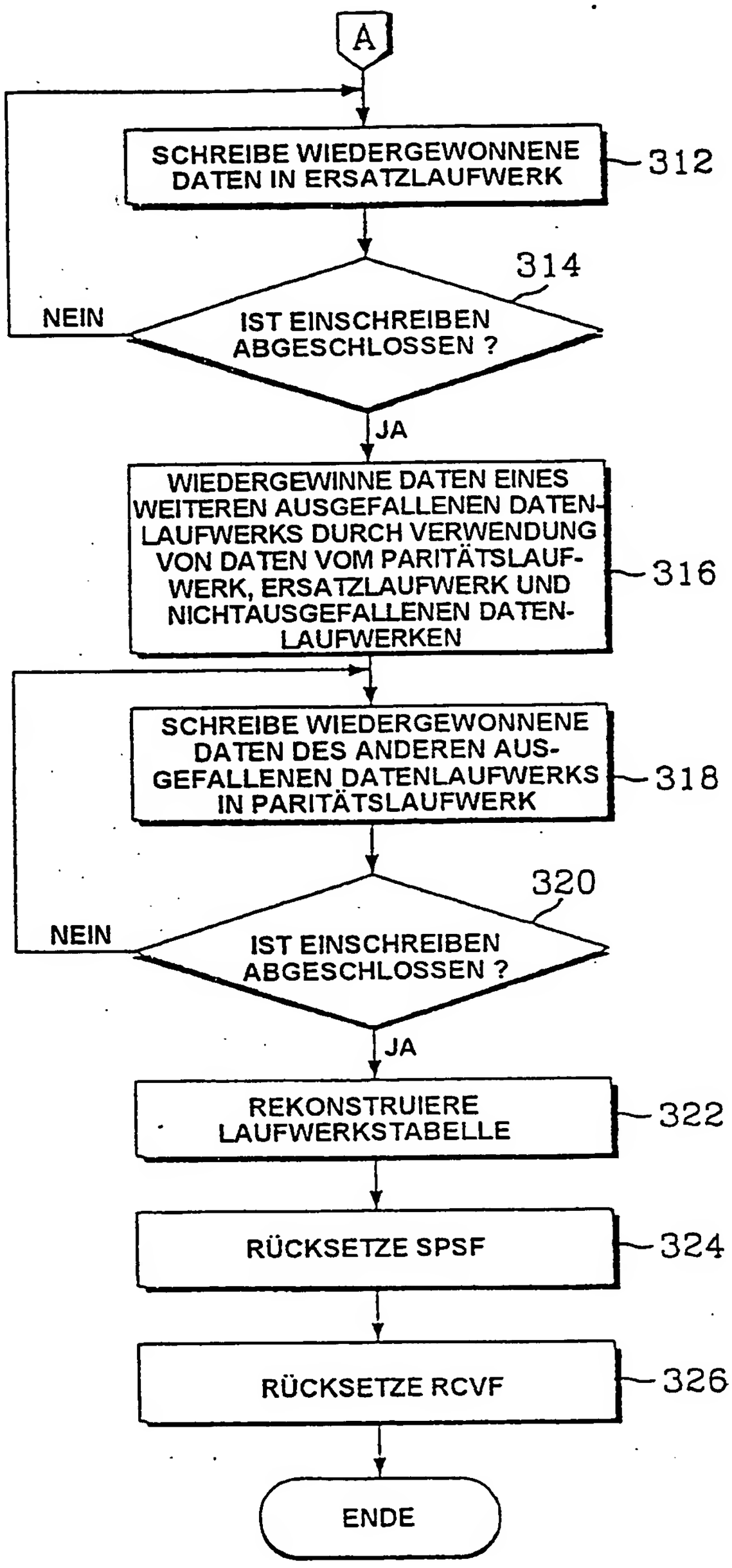


Fig. 10B

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

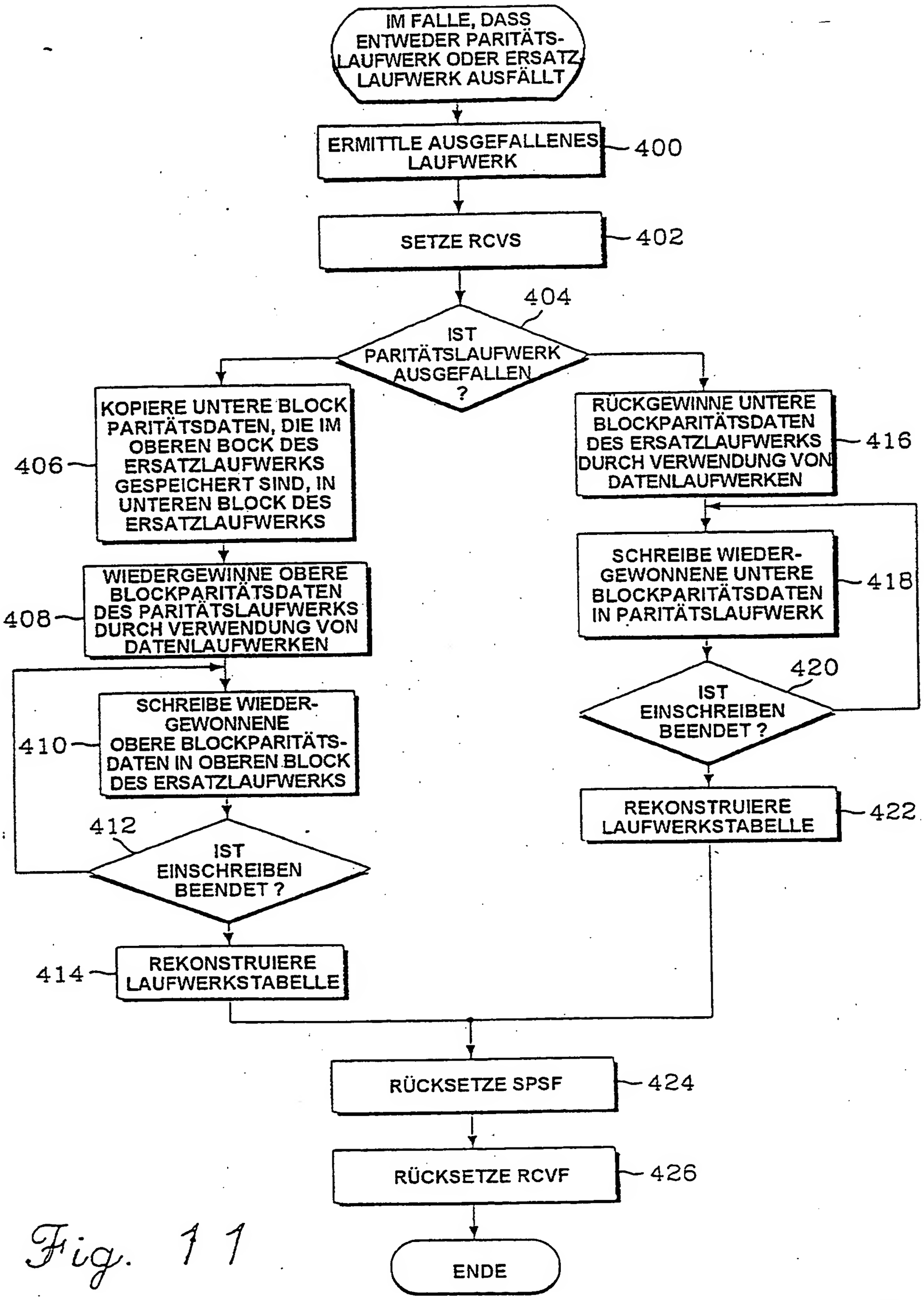


Fig. 11

802 014/650

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenl gungstag: 2. April 1998

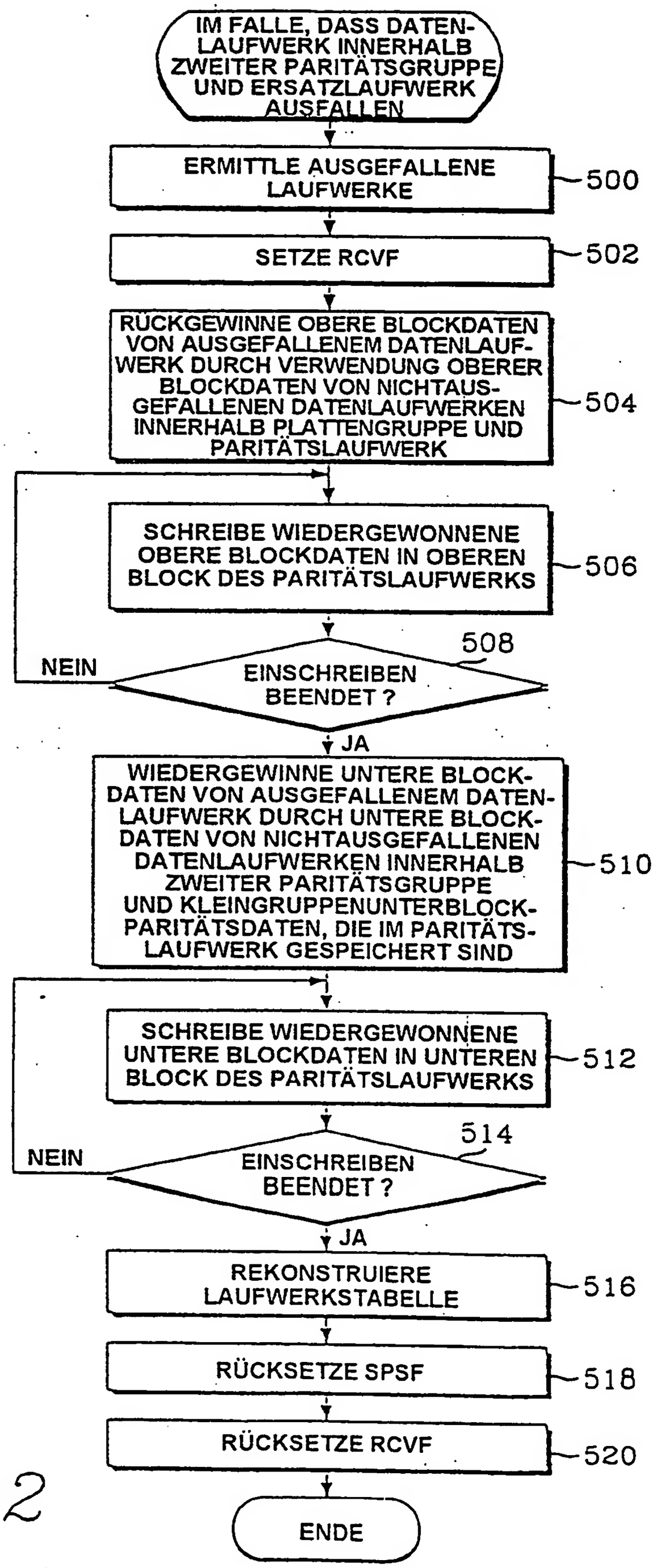
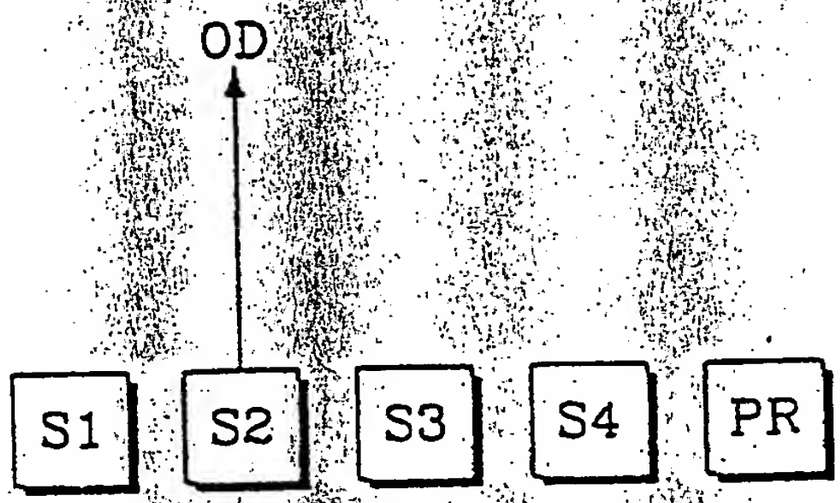


Fig. 12

802 014/650

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

(A)



(B)

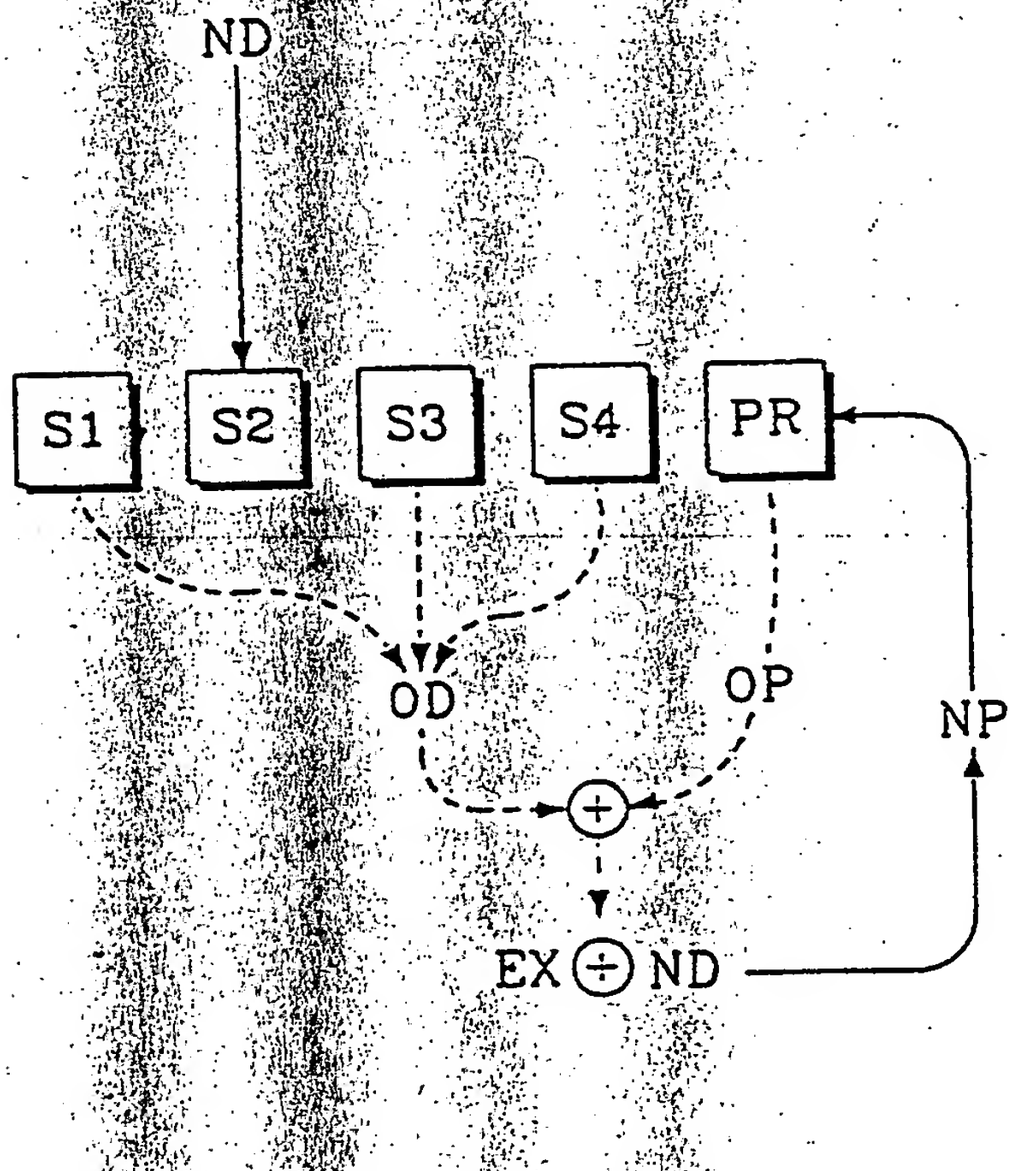


Fig. 13

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

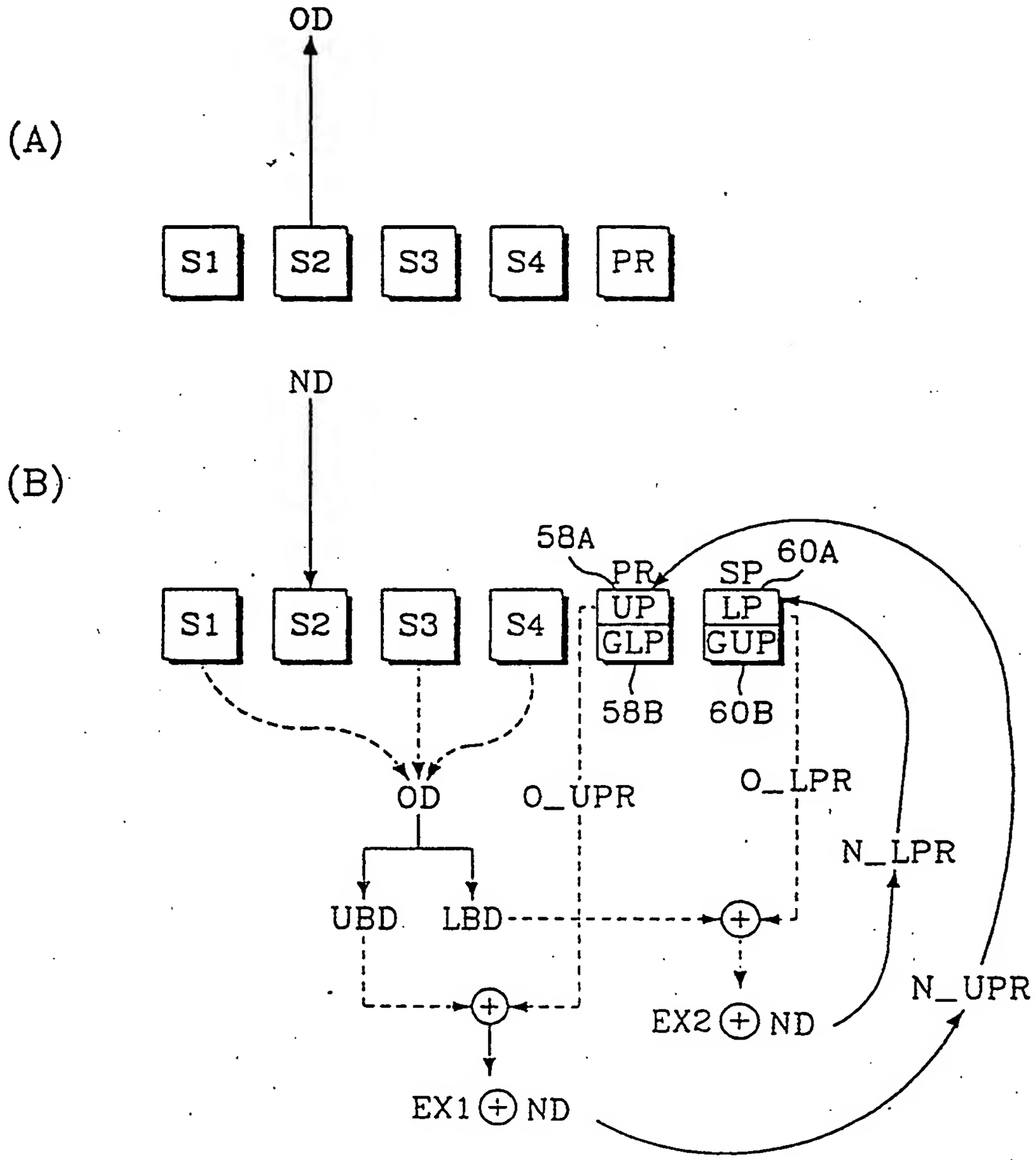


Fig. 14

Nummer: DE 197 23 909 A1
Int. Cl.[6]: G 11 B 20/18
Offenlegungstag: 2. April 1998

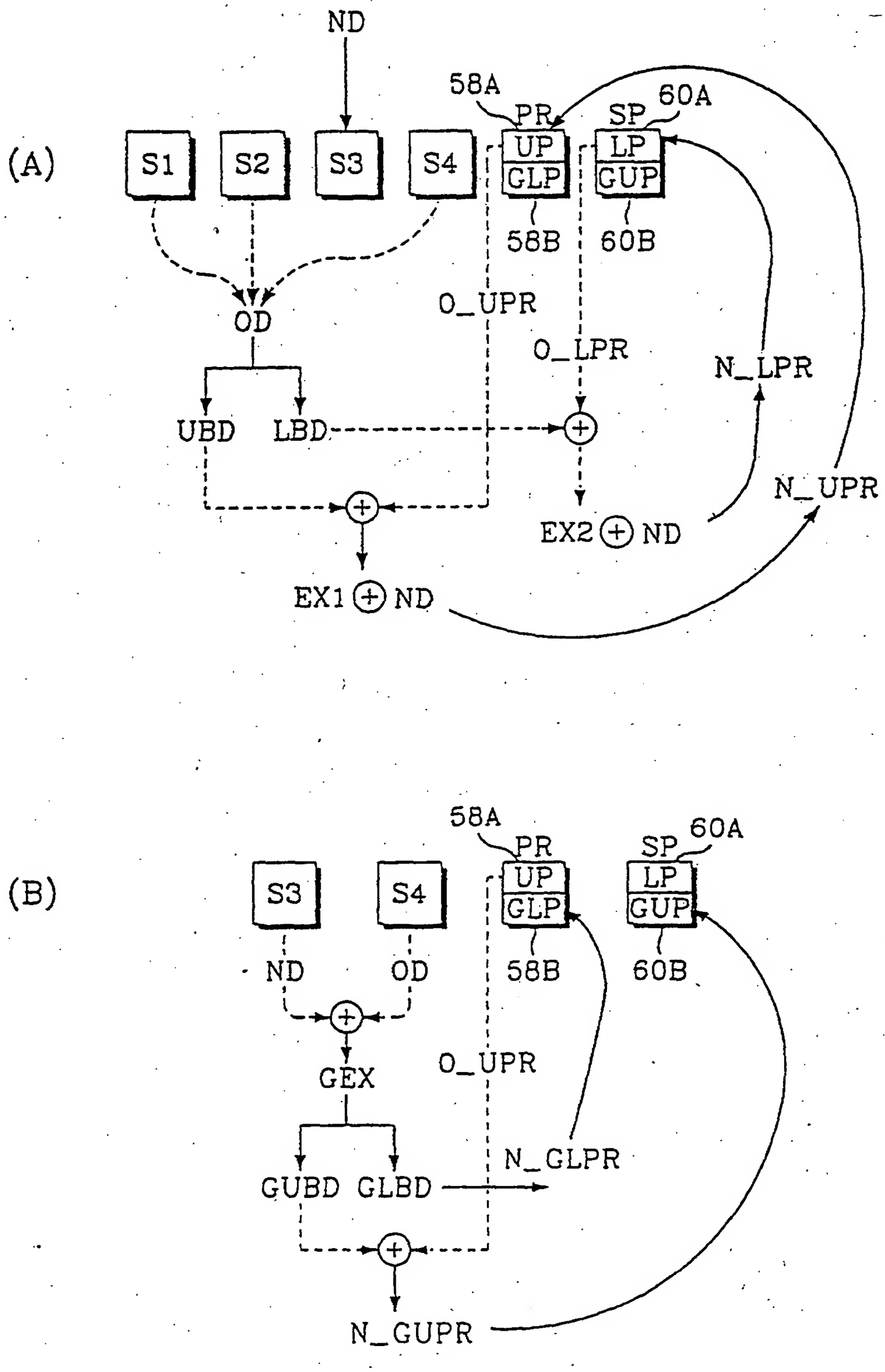


Fig. 15